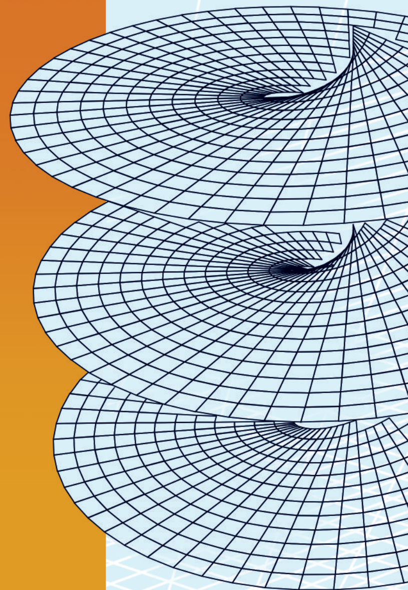
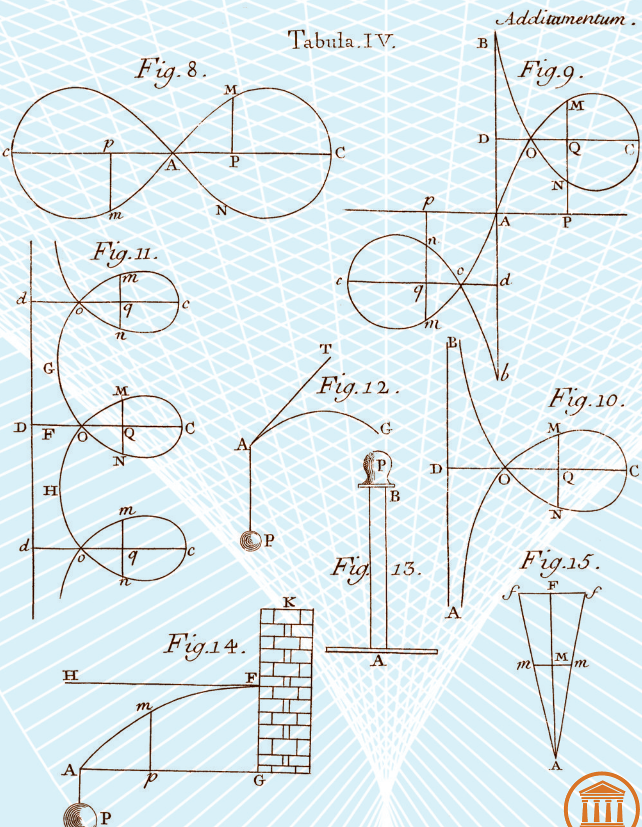
STUDENT MATHEMATICAL LIBRARY
Volume 72

# A First Course in the Calculus of Variations

Mark Kot






AMS
AMERICAN MATHEMATICAL SOCIETY

# A First Course in the Calculus of Variations

STUDENT MATHEMATICAL LIBRARY
Volume 72

# A First Course in the Calculus of Variations

Mark Kot

American Mathematical Society
Providence, Rhode Island

2000 *Mathematics Subject Classification*. Primary 49-01.

For additional information and updates on this book, visit
**www.ams.org/bookpages/stml-72**

**Library of Congress Cataloging-in-Publication Data**

Kot, Mark, 1956–
A first course in the calculus of variations / Mark Kot.
pages cm. — (Student mathematical library ; volume 72)
Includes bibliographical references and index.
ISBN 978-1-4704-1495-5 (alk. paper)
1. Calculus of variations—Textbooks 2. Calculus of variations—Study and teaching (Higher) I. Title.

QA315.K744 2014
515′.64—dc23

2014024014



∞ The paper used in this book is acid-free and falls within the guidelines established to ensure permanence and durability.
Visit the AMS home page at `http://www.ams.org/`

10 9 8 7 6 5 4 3 2 1 19 18 17 16 15 14

# Contents

# Preface

This book is intended for a first course, at the senior or beginning graduate level, in the calculus of variations. It will also be of use to those interested in self-study.

There are already many excellent books on this topic. I cite a number of these texts throughout this book. I have added another book, this book, because I wanted a text that is especially well suited to the Amath 507 class that I teach at the University of Washington.

My Amath 507 students are typically applied mathematicians, physicists, and engineers. I have thus included numerous examples from fields such as mechanics and optics; I have also included many examples with immediate geometric appeal. Because of my students' strong interest in applications, I have also introduced constraints earlier than usual.

My students also enjoy learning the history of science. So I have resisted the temptation of immediately jumping to the most modern results. I instead follow the historical development of the calculus of variations. The calculus of variations has an especially rich and interesting history and a historical approach works exceptionally well for this subject.

Finally, I teach on a quarter system. So I have taken the opportunity of writing this book to collect and organize my thoughts on the calculus of variations in what I hope is a concise and effective manner.

I am grateful to my Amath 507 students for their enthusiasm and hard work and for uncovering interesting applications of the calculus of variations. I owe special thanks to William K. Smith for supervising my undergraduate thesis in the calculus of variations (35 years ago) and to Hanno Rund for teaching a fine series of courses on the calculus of variations during my graduate years. Sadly, these two wonderful teachers are now both deceased. I thank the fine editors and reviewers of the American Mathematical Society for their helpful comments. Finally, I thank my family for their encouragement and for putting up with the writing of another book.

Mark Kot

Chapter 1

# Introduction

## 1.1. The brachistochrone

The calculus of variations has a clear starting point. In June of 1696, John (also known as Johann or Jean) Bernoulli challenged the greatest mathematicians of the world to solve the following new problem (Bernoulli, 1696; Goldstine, 1980):

> Given points $A$ and $B$ in a vertical plane to find the path $AMB$ down which a movable point $M$ must, by virtue of its weight, proceed from $A$ to $B$ in the shortest possible time.

Imagine a particle $M$ of mass $m$, in a vertical gravitational field of strength $g$, that moves along the curve $y = y(x)$ between the two points $A = (a, y_a)$ and $B = (b, y_b)$ (see Figure 1.1). The time of descent $T$ of the particle is

$$T = \int_0^T dt = \int_0^L \frac{dt}{ds}\, ds = \int_0^L \frac{1}{v}\, ds = \int_a^b \frac{1}{v} \sqrt{1 + y'^2}\, dx\,, \qquad (1.1)$$

where $s$ is arc length, $L$ is the length of the curve, and $v$ is the speed of the particle.

If our particle moves without friction, the law of conservation of mechanical energy guarantees that the sum of the particle's kinetic

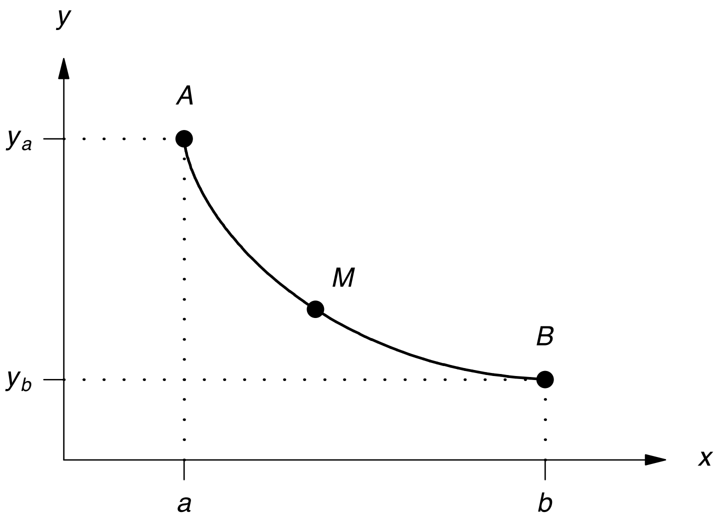



**Figure 1.1.** Curve of descent

energy and potential energy remains constant. If our particle starts from rest, we may thus write

$$\frac{1}{2}mv^2 + mgy \;=\; mgy_a\,. \tag{1.2}$$

The particle's speed is then

$$v \;=\; \sqrt{2g(y_a - y)}\,. \tag{1.3}$$

We now wish to find the *brachistochrone* (from $\beta\rho\alpha\chi\iota\sigma\tau o\varsigma$, shortest, and $\chi\rho o\nu o\varsigma$, time; John Bernoulli originally, but erroneously, wrote brachystochrone). That is, we wish to find the curve

$$y \;=\; y(x) \;\leq\; y_a \tag{1.4}$$

that minimizes the integral

$$T \;=\; \frac{1}{\sqrt{2g}} \int_a^b \sqrt{\frac{1 + y'^2}{y_a - y}}\; dx\,. \tag{1.5}$$

Several famous mathematicians responded to John Bernoulli's challenge. Solutions were submitted by Gottfried Wilhelm Leibniz

(1697), Isaac Newton (1695–7, 1697), John Bernoulli (1697a), James (or Jakob) Bernoulli (1697), and Guillaume l'Hôpital (1697).

Leibniz provided a geometrical solution. He derived the differential equation for the brachistochrone but did not specify the resulting curve (Goldstine, 1980). Leibniz also suggested that the brachistochrone be called the *tachystoptotam* (from $\tau\alpha\chi\iota\sigma\tau o\varsigma$, swiftest, and $\pi\iota\pi\tau\epsilon\iota\nu$, to fall). Mercifully, this suggestion was ignored.

Newton's anonymous solution was published in the *Philosophical Transactions*; it was then reprinted in the *Acta Eruditorum*. Newton provided the correct answer but gave no clue to his method. Despite Newton's anonymity, John Bernoulli recognized that the work was "ex ungue Leonem" (from the claw of the Lion) and the *Acta Eruditorum* listed Newton in its index of authors.

John Bernoulli provided two solutions. The first solution relied on an analogy between the mechanical brachistochrone and light. Bernoulli (1697a) was quite taken with Fermat's principle of least time for light and argued that the brachistochrone "is the curve that a light ray would follow on its way through a medium whose density is inversely proportional to the velocity that a heavy body acquires during its fall." He broke up the optical medium into thin horizontal layers, chose an appropriate index of refraction, and used Snell's law of refraction and calculus to determine the shape of the brachistochrone. John Bernoulli (1718) described his second solution many years later. This second solution received little attention at the time but is now viewed as the first sufficiency proof in the calculus of variations.

James Bernoulli's solution was not as elegant as that of his younger brother, but it contained the key idea of varying only one value of the solution curve at a time. This idea provided the basis for further work in the calculus of variations. James Bernoulli called his solution an *oligochrone* (from $o\lambda\iota\gamma o\varsigma$, little, and $\chi\rho o\nu o\varsigma$, time).

We shall see that the brachistochrone is the inverted cycloid

$$x(\phi) \;=\; a + R(\phi - \sin\,\phi)\,, \quad y(\phi) \;=\; y_a - R\,(1 - \cos\,\phi)\,, \qquad (1.6)$$

where the parameter $R$ is uniquely determined by the initial and terminal points. This cycloid is the curve traced by a point on the

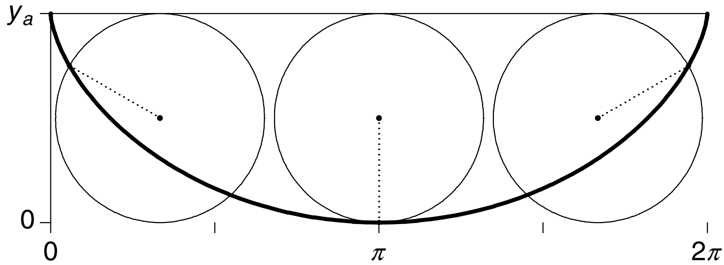



**Figure 1.2.** Cycloid for $R = \frac{1}{2} y_a$ and $a = 0$

circumference of a circle of radius $R$ rolling along the bottom of the horizontal line $y = y_a$ (see Figure 1.2).

Huygens (1673, 1986) had previously shown that an inverted cycloid is a tautochrone (from $\tau\alpha\upsilon\tau o$ or $\tau o\ \alpha\upsilon\tau o$, the same, and $\chi\rho o\nu o\varsigma$, time): the time for a heavy particle to fall to the bottom of this curve is independent of the upper starting point. To John Bernoulli's astonishment, the brachistochrone was Huygens' tautochrone.

The brachistochrone is one of many problems where we wish to determine a function, $y(x)$, that minimizes or maximizes the integral

$$J[y(x)] \;=\; \int_a^b f(x, y(x), y'(x))\; dx\,. \tag{1.7}$$

Leonhard Euler first devised a systematic method for solving such problems.

In the remainder of this chapter, we will examine three other problems that involve minimizing or maximizing integrals. We will first look at another brachistochrone problem, for travel *through* the earth. We will then look at the problem of finding the shortest path between two points on some general surface. Finally, we will look at the "soap-film problem," the problem of minimizing the surface area of a surface of revolution. All of these problems can be attacked using the calculus of variations.

## 1.2. The terrestrial brachistochrone

History repeats itself. In August of 1965, *Scientific American* published an article on "High-Speed Tube Transportation" (Edwards, 1965). Edwards proposed tube trains that would fall through the earth, pulled by gravity and helped along by pneumatic propulsion. The advantages cited by Edwards included:

> (1) It brings most of the tunnel down into deep bedrock, where the cost of tunneling — by blasting or by boring — is reduced and incidental earth shifts are minimized; the rock is more homogeneous in consistency and there is less likelihood of water inflow.
>
> (2) The nuisance to property owners decreases with depth, so the cost of easements should be lower.
>
> (3) A deep tunnel does not interfere with subways, building foundations, utilities, or water wells....
>
> (4) The pendulum ride is uniquely comfortable for the passenger....

Lest you think this pure fantasy, a pneumatic train was constructed in New York City, under Broadway, from Warren Street to Murray Street, in 1870 by Alfred Ely Beach (an early owner of *Scientific American*). This was New York City's first subway (Roess and Sansone, 2013). You can see a drawing of the pneumatic train on the wallpaper in older Subway Sandwich shops.

Cooper (1966a) then pointed out that straight-line chords lead to needlessly long trips through the earth. He used the calculus of variations to derive a differential equation for the fastest tunnels through the earth and integrated this equation numerically. Venezian (1966), Mallett (1966), Laslett (1966), and Patel (1967) then found first integrals and analytic solutions for this problem. See Cooper (1966b) for a summary.

Let us take a closer look at this *terrestrial brachistochrone* problem. Assume that the earth is a homogeneous sphere of radius $R$. Consider a section through the earth with polar coordinates centered at the heart of the earth (see Figure 1.3). Imagine a particle of mass

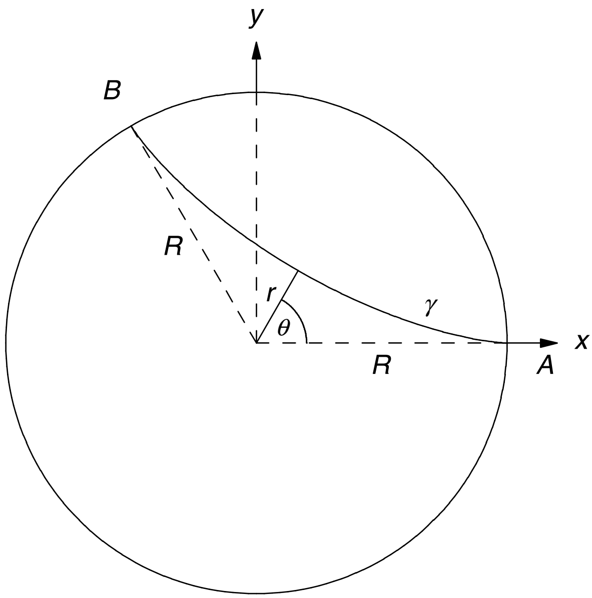



**Figure 1.3.** Path through the earth

$m$ that moves between two points, $A = (r_a, \theta_a)$ and $B = (r_b, \theta_b)$, on or near the surface of the earth. We now wish to find the planar curve $\gamma$ that minimizes the travel time

$$T = \int_0^T dt = \int_\gamma \frac{dt}{ds}\, ds = \int_\gamma \frac{1}{v}\, ds = \int_\gamma \frac{1}{v} \sqrt{dr^2 + r^2\, d\theta^2} \qquad (1.8)$$

between $A$ and $B$, where $s$ is arc length and $v$ is the speed of the particle.

When a particle is outside a uniform spherical shell, the shell exerts a gravitational force equal to that of an identical point mass at the center of the shell. A particle inside the shell feels no force (see Exercise 1.6.3). By integrating over spherical shells of different radii (Exercise 1.6.4), one can show that the gravitational potential energy within a spherical and homogeneous earth can be written

$$V(r) = \frac{1}{2}\frac{mg}{R} r^2\,, \qquad (1.9)$$

where $g$ is the magnitude of the gravitational acceleration at the surface of the earth.

For a particle starting at rest at the surface of the earth, conservation of energy now implies that

$$\frac{1}{2}mv^2 + \frac{1}{2}\frac{mg}{R}r^2 \;=\; \frac{1}{2}mgR \tag{1.10}$$

so that

$$v \;=\; \frac{\sqrt{g(R^2 - r^2)}}{\sqrt{R}}\,. \tag{1.11}$$

It follows that the total travel time is

$$T \;=\; \sqrt{\frac{R}{g}} \int_{\theta_a}^{\theta_b} \sqrt{\frac{\left(\frac{dr}{d\theta}\right)^2 + r^2}{R^2 - r^2}}\, d\theta\,. \tag{1.12}$$

We will look at this problem in greater detail later. We shall see that the terrestrial brachistochrone is a *hypocycloid,* the curve traced by a point on the circumference of a circle of radius either $[R - (S_{AB}/\pi)]$ (see Figure 1.4) or of radius $S_{AB}/\pi$ (see Figure 1.5), where $S_{AB}$ is the arc length along the surface of the earth between $A$ and $B$, as it rolls inside a circle of radius $R$.

The fastest Amtrak train makes the 400 mile trip between Boston and Washington, D.C., in six and a half hours. A tube train moving along a straight-line chord between Boston and Washington would penetrate 5 miles into the earth and take 42 minutes. The fastest tube train along a hypocycloid would, in turn, penetrate 125 miles into the earth and take 10.7 minutes.

## 1.3. Geodesics

I do not want to give the impression that the calculus of variations is only brachistochrones. In this and the next section, we will look at two other classic problems.

A line is the shortest path between two points in a plane. We also wish to find shortest paths between pairs of points on other, more general, surfaces. To find these geodesics, we must minimize arc length.

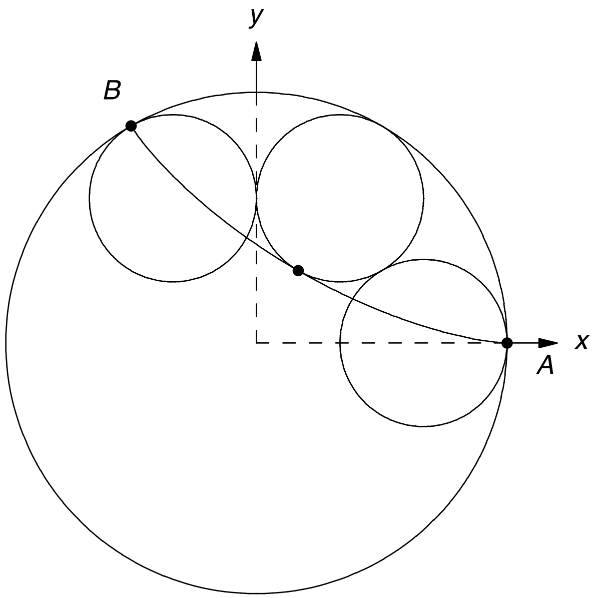



**Figure 1.4.** Hypocycloid with inner radius $\left(R - \frac{S_{AB}}{\pi}\right)$

The simplest case arises when the surface is a level set for one of the coordinates in a system of orthogonal curvilinear coordinates. The arc length can then be written using the scale factors of the coordinate system.

Consider, for example, two points, $A$ and $B$, on a sphere of radius $R$ centered at the origin. We wish to join $A$ and $B$ by the shortest, continuously differentiable curve lying on the sphere. We start by specifying position,

$$\mathbf{r}(x, y, z) \;=\; x\,\mathbf{i} \;+\; y\,\mathbf{j} \;+\; z\,\mathbf{k}\,, \tag{1.13}$$

using the Cartesian coordinates $x$, $y$, and $z$ and Cartesian basis vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$. For points on the surface of a sphere, we now switch to the spherical coordinates $r$, $\theta$, and $\phi$ (see Figure 1.6). Since

$$x = r\sin\theta\cos\phi\,, \quad y = r\sin\theta\sin\phi\,, \quad z = r\cos\theta\,, \tag{1.14}$$

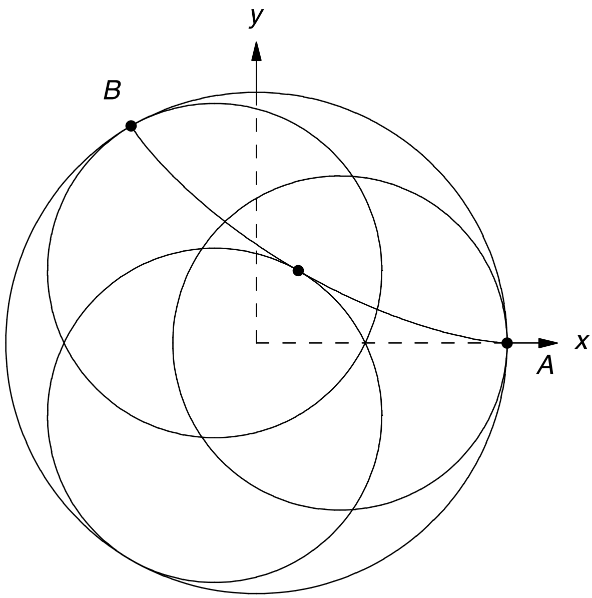



**Figure 1.5.** Hypocycloid with inner radius $\frac{S_{AB}}{\pi}$

the position vector **r** now takes the form

$$\mathbf{r}(r,\theta,\phi) \;=\; r\,\sin\,\theta\,\cos\,\phi\;\mathbf{i} + r\,\sin\,\theta\,\sin\,\phi\;\mathbf{j} + r\,\cos\,\theta\;\mathbf{k}\,. \quad (1.15)$$

Since this position vector depends on $r$, $\theta$, and $\phi$,

$$d\mathbf{r} \;=\; \frac{\partial \mathbf{r}}{\partial r}\,dr + \frac{\partial \mathbf{r}}{\partial \theta}\,d\theta + \frac{\partial \mathbf{r}}{\partial \phi}\,d\phi\,. \quad (1.16)$$

The three partial derivatives on the right-hand side of this equation are vectors tangent to motions in the $r$, $\theta$, and $\phi$ directions. Thus

$$d\mathbf{r} \;=\; h_r\,dr\,\hat{e}_r + h_\theta\,d\theta\,\hat{e}_\theta + h_\phi\,d\phi\,\hat{e}_\phi\,, \quad (1.17)$$

where $\hat{e}_r$, $\hat{e}_\theta$, and $\hat{e}_\phi$ are unit vectors in the $r$, $\theta$, and $\phi$ directions and

$$h_r = \left|\left|\frac{\partial \mathbf{r}}{\partial r}\right|\right| = 1\,,\;\; h_\theta = \left|\left|\frac{\partial \mathbf{r}}{\partial \theta}\right|\right| = r\,,\;\; h_\phi = \left|\left|\frac{\partial \mathbf{r}}{\partial \phi}\right|\right| = r\,\sin\,\theta \quad (1.18)$$

are the scale factors for spherical coordinates.

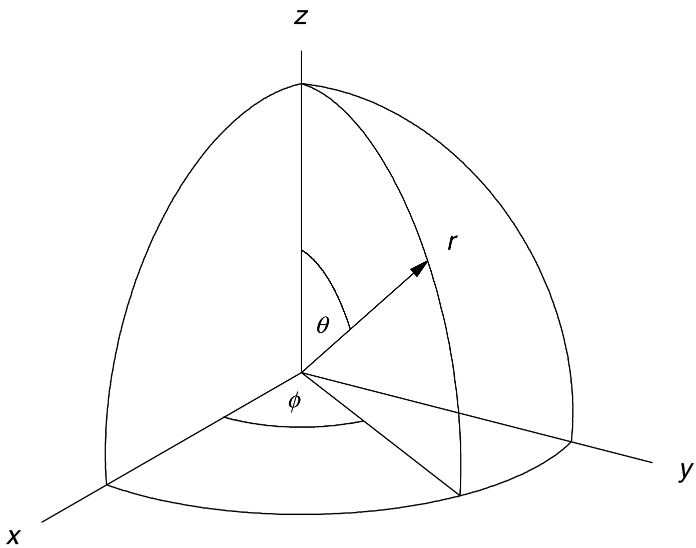



**Figure 1.6.** Spherical coordinates

The element of arc length in spherical coordinates is given by

$$ds = \sqrt{d\mathbf{r}\cdot d\mathbf{r}} = \sqrt{h_r^2\,dr^2 + h_\theta^2\,d\theta^2 + h_\phi^2\,d\phi^2} \qquad (1.19)$$
$$= \sqrt{dr^2 + r^2\,d\theta^2 + r^2\,\sin^2\theta\,d\phi^2}\,.$$

For a sphere of radius $r = R$, this element reduces to

$$ds = R\sqrt{d\theta^2 + \sin^2\theta\,d\phi^2}\,. \qquad (1.20)$$

If we assume that $\phi = \phi(\theta)$, finding the curve that minimizes the arc length between the points $A = (\theta_a, \phi_a)$ and $B = (\theta_b, \phi_b)$ simplifies to finding the function $\phi(\theta)$ that minimizes the integral

$$s = \int_A^B ds = R\int_{\theta_A}^{\theta_B} \sqrt{1 + \sin^2\theta\,(d\phi/d\theta)^2}\;d\theta \qquad (1.21)$$

subject to the boundary conditions

$$\phi(\theta_a) = \phi_a\,, \quad \phi(\theta_b) = \phi_b\,. \qquad (1.22)$$

We will see, later, that the shortest paths on a sphere are arcs of great circles.

Unfortunately, we cannot expect every interesting surface to be the level set for some common coordinate. We may, however, hope to represent our surface parametrically. We may prescribe the $x$, $y$, and $z$ coordinates of points on the surface using the parameters $u$ and $v$ and write our surface in the vector form

$$\mathbf{r}(u,v) \;=\; x(u,v)\,\mathbf{i} \;+\; y(u,v)\,\mathbf{j} \;+\; z(u,v)\,\mathbf{k}\,. \tag{1.23}$$

We can now specify a curve on this surface by prescribing $u$ and $v$ in terms of a single parameter — call it $t$ — so that

$$u \;=\; u(t)\,, \quad v \;=\; v(t)\,. \tag{1.24}$$

The vector

$$\dot{\mathbf{r}} \;\equiv\; \frac{d\mathbf{r}}{dt} \;=\; \frac{\partial \mathbf{r}}{\partial u}\,\dot{u} + \frac{\partial \mathbf{r}}{\partial v}\,\dot{v} \tag{1.25}$$

is tangent to both the curve and the surface. We find the square of the distance between two points on a curve by integrating

$$ds^2 \;=\; d\mathbf{r}\cdot d\mathbf{r} \;=\; \left(\frac{\partial \mathbf{r}}{\partial u}\,du + \frac{\partial \mathbf{r}}{\partial v}\,dv\right)\cdot\left(\frac{\partial \mathbf{r}}{\partial u}\,du + \frac{\partial \mathbf{r}}{\partial v}\,dv\right) \tag{1.26}$$

along the curve. Equation (1.26) is often written

$$ds^2 \;=\; E\,du^2 + 2\,F\,du\,dv + G\,dv^2\,, \tag{1.27}$$

where

$$E \;=\; \frac{\partial \mathbf{r}}{\partial u}\cdot\frac{\partial \mathbf{r}}{\partial u}\,, \quad F \;=\; \frac{\partial \mathbf{r}}{\partial u}\cdot\frac{\partial \mathbf{r}}{\partial v}\,, \quad G \;=\; \frac{\partial \mathbf{r}}{\partial v}\cdot\frac{\partial \mathbf{r}}{\partial v}\,. \tag{1.28}$$

The right-hand side of equation (1.27) is called the *first fundamental form* of the surface. The coefficients $E(u,v)$, $F(u,v)$, and $G(u,v)$ have many names. They are sometimes called first-order fundamental magnitudes or quantities. Other times, they are simply called the coefficients of the first fundamental form.

The distance between the two points $A = (u_a, v_a)$ and $B = (u_b, v_b)$ on the curve $u = u(t)$, $v = v(t)$ may now be written

$$s \;=\; \int_{t_a}^{t_b} \sqrt{E\left(\frac{du}{dt}\right)^2 + 2\,F\frac{du}{dt}\frac{dv}{dt} + G\left(\frac{dv}{dt}\right)^2}\;dt\,, \tag{1.29}$$

with

$$u(t_a) = u_a\,, \quad v(t_a) = v_a\,, \quad u(t_b) = u_b\,, \quad v(t_b) = v_b\,. \tag{1.30}$$

In this formulation, we have two dependent variables, $u(t)$ and $v(t)$, and one independent variable, $t$. If $v$ can be written as a function of $u$, $v = v(u)$, we can instead rewrite our integral as

$$s = \int_{u_a}^{u_b} \sqrt{E + 2\,F\left(\frac{dv}{du}\right) + G\left(\frac{dv}{du}\right)^2}\; du \tag{1.31}$$

with

$$v(u_a) = v_a, \quad v(u_b) = v_b\,. \tag{1.32}$$

This is now a problem with one dependent variable and one independent variable.

To make all this concrete, let us take, as an example, the *pseudosphere* (see Figure 1.7), half of the surface of revolution generated by rotating a tractrix about its asymptote. If the asymptote is the $z$-axis, we can write the equation for a pseudosphere, parametrically, as

$$\begin{aligned} \mathbf{r}(u,v) &= a \sin u \cos v\; \mathbf{i} \;+\; a \sin u \sin v\; \mathbf{j} \\ &\quad + a\left(\cos u + \ln \tan \frac{u}{2}\right)\; \mathbf{k}\,. \end{aligned} \tag{1.33}$$

Since

$$\begin{aligned} \mathbf{r}_u &= \frac{\partial \mathbf{r}}{\partial u} \\ &= (a \cos u \cos v,\, a \cos u \sin v, -a \sin u + a \csc u) \end{aligned} \tag{1.34}$$

and

$$\mathbf{r}_v = \frac{\partial \mathbf{r}}{\partial v} = (-a \sin u \sin v,\, a \sin u \cos v,\, 0)\,, \tag{1.35}$$

the first-order fundamental quantities reduce to

$$E = \mathbf{r}_u \cdot \mathbf{r}_u = a^2 \cot^2 u\,, \tag{1.36}$$

$$F = \mathbf{r}_u \cdot \mathbf{r}_v = 0\,, \tag{1.37}$$

$$G = \mathbf{r}_v \cdot \mathbf{r}_v = a^2 \sin^2 u\,. \tag{1.38}$$

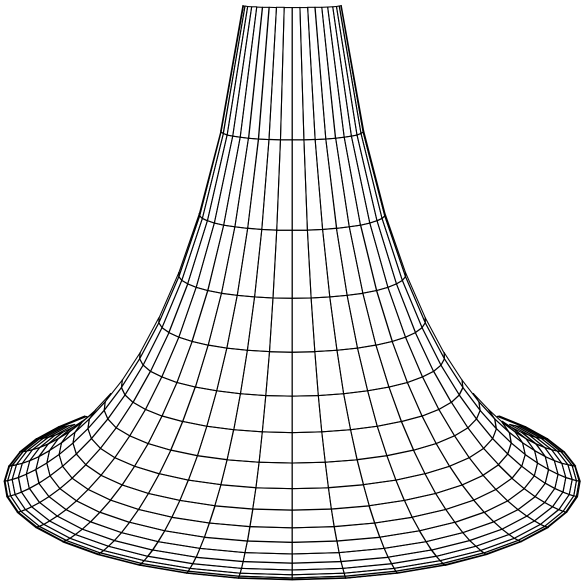


**Figure 1.7.** Pseudosphere

To determine a geodesic on the pseudosphere, we must thus find a curve, $u = u(t)$ and $v = v(t)$, that minimizes the arc-length integral

$$s \;=\; a \int_{t_a}^{t_b} \sqrt{\cot^2 u\, \dot{u}^2 + \sin^2 u\, \dot{v}^2}\; dt \tag{1.39}$$

subject to the boundary conditions

$$u(t_a) \;=\; u_a\,, \quad v(t_a) \;=\; v_a\,, \quad u(t_b) \;=\; u_b\,, \quad v(t_b) \;=\; v_b\,. \tag{1.40}$$

Alternatively, we may look for a curve, $v = v(u)$, that minimizes the integral

$$s \;=\; a \int_{u_a}^{u_b} \sqrt{\cot^2 u + \sin^2 u \left(\frac{dv}{du}\right)^2}\; du \tag{1.41}$$

subject to the boundary conditions

$$v(u_a) \;=\; v_a, \quad v(u_b) \;=\; v_b\,. \tag{1.42}$$

For other examples, see Exercise 1.6.6.

John Bernoulli (1697b) posed the problem of finding geodesics on convex surfaces. In 1698, he remarked, in a letter to Leibniz, that geodesics always have osculating planes that cut the surface at right angles. (An osculating plane is the plane that passes through three nearby points on a curve as two of these points approach the third point.) This geometric property is frequently used as the definition of a geodesic curve, irrespective of whether the curve actually minimizes arc length. Later, Euler (1732) derived differential equations for geodesics on surfaces using the calculus of variations. This was Euler's earliest known use of the calculus of variations.

Finding shortest paths is easiest on simple surfaces of revolution. Geodesics on surfaces of revolution satisfy a simple first integral or "conservation law" that was first published by Clairaut (1733). Jacobi (1839), in a tour de force, succeeded in integrating the equations of geodesics for a more complicated surface, a triaxial ellipsoid.

## 1.4. Minimal surfaces

We may minimize areas as well as lengths. Consider two points,

$$y(a) = y_a\,, \quad y(b) = y_b\,, \tag{1.43}$$

in the plane (see Figure 1.8). We wish to join these two points by a continuously differentiable curve,

$$y = y(x) \geq 0\,, \tag{1.44}$$

in such a way that the surface of revolution, generated by rotating this curve about the $x$-axis, has the smallest possible area $S$. In other words, we wish to minimize

$$S = 2\pi \int_a^b y(x)\sqrt{1 + y'^2}\,dx\,. \tag{1.45}$$

Some of you will recognize this as the "soap-film problem." Suppose we wish to find the shape of a soap film that connects two wire hoops. For a soap film with constant film tension, the surface energy is proportional to the area of the film. Minimizing the surface energy of the film is thus equivalent to minimizing its surface area (Isenberg,

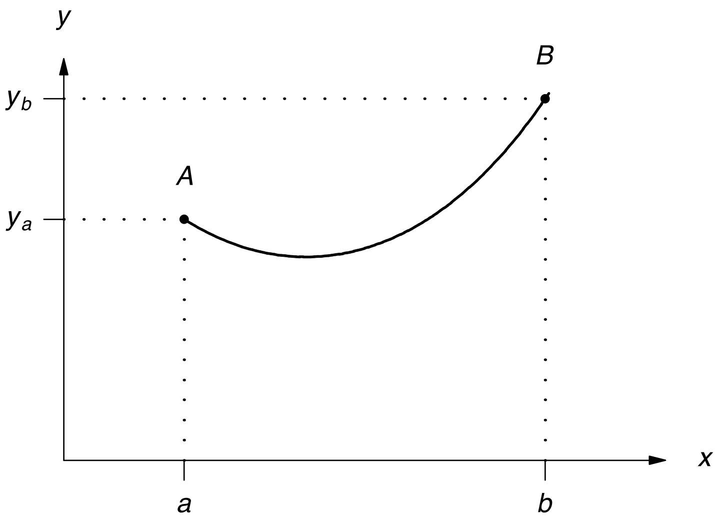



**Figure 1.8.** Profile curve

1992; Oprea, 2000). (For a closed soap *bubble,* without fixed boundaries, excess air pressure within the bubble prevents the surface area of the bubble from shrinking to zero.)

Euler (1744) discovered that the *catenoid,* the surface generated by a catenary or hanging chain (see Figure 1.9), minimizes surface area. As you doubtless know, however, from playing with soap films, if you pull two parallel hoops too far apart, the catenoid breaks, leaving soap film on the hoops. This was first shown analytically by Goldschmidt (1831). For two parallel, coaxial hoops of radius $r$, the area of a catenoid is an absolute minimum if the distance between the hoops is less than $1.056\,r$. This area is a relative minimum for distances between $1.056\,r$ and $1.325\,r$. For distances greater than $1.325\,r$, the catenoid breaks and the solution jumps to the discontinuous *Goldschmidt solution* (two disks).

Joseph Lagrange (1762) then proposed the general problem of finding a surface, $z = f(x, y)$, with a closed curve $C$ as its boundary, that has the smallest area. That is, we now wish to minimize a *double*

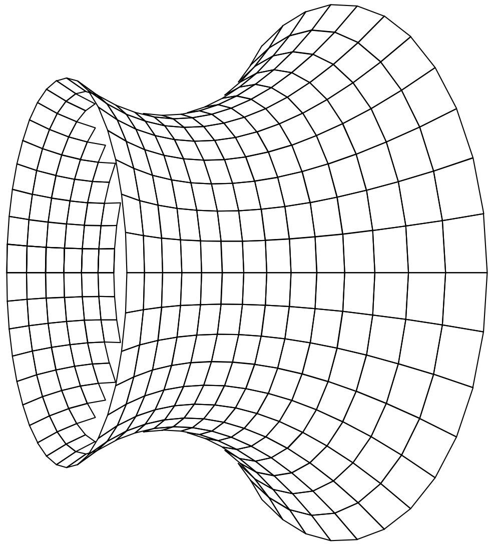


**Figure 1.9.** Catenoid

integral of the form

$$S = \iint_{\Omega} \sqrt{1 + f_x^2 + f_y^2}\, dx\, dy \qquad (1.46)$$

(see Exercise 1.6.7), where $\partial\Omega$ is the projection of the closed curve $C$ onto the $(x, y)$ plane and $\Omega$ is the interior of this projection. This problem has been known, starting with Lebesgue (1902), as Plateau's problem, in honor of Joseph Plateau's extensive experiments (Plateau, 1873) with soap films.

Lagrange showed that a surface that minimizes integral (1.46) must satisfy the *minimal surface equation*

$$(1 + f_y^2)\, f_{xx} - 2\, f_x\, f_y\, f_{xy} + (1 + f_x^2)\, f_{yy} = 0\,, \qquad (1.47)$$

a quasilinear, elliptic, second-order, partial differential equation. Different constraints on the function $f(x, y)$ (e.g., Exercise 1.6.10) yield different minimal surfaces.

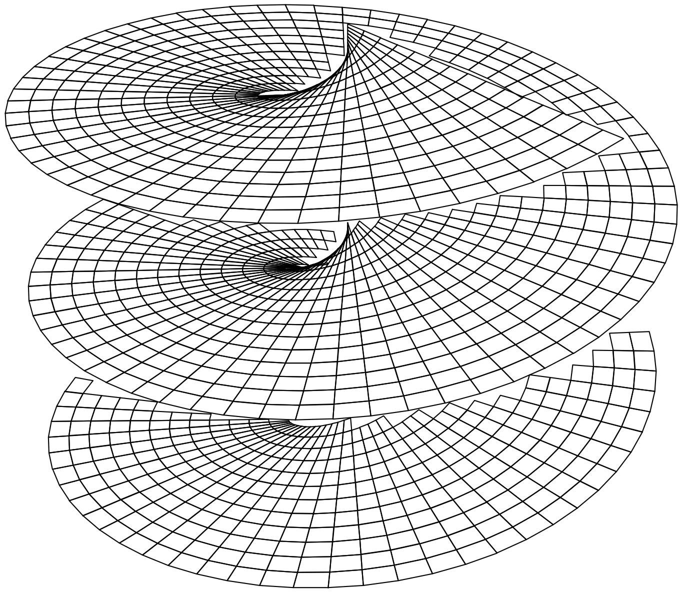


**Figure 1.10.** Helicoid

Jean-Baptiste-Marie-Charles Meusnier (1785) soon gave equation (1.47) a geometric interpretation. At each point $P$ of a smooth surface, choose a vector normal to the surface, cut the surface with normal planes (that contain the normal vector but that differ in orientation), and obtain a series of plane curves. For each plane curve, determine the curvature at $P$. Find the minimum and maximum curvatures (from amongst all the plane curves passing through $P$). These are your *principal curvatures*.

Meusnier showed that the minimal surface equation implies that the *mean curvature* (the average of the principal curvatures) is zero at every point of the minimizing surface. As a result, any surface with zero mean curvature is typically referred to as a minimal surface, even if it does not provide an absolute or relative minimum for surface area. Meusnier also discovered that the catenoid and the *helicoid*, the surface formed by line segments perpendicular to the axis of a circular helix as they go through the helix (see Figure 1.10), satisfy Lagrange's minimal surface equation. (Meusnier, like Lagrange, seemed unaware of Euler's earlier analysis of the catenoid.) The study of minimal surfaces has grown to become one of the richest areas of mathematical research.

In the remainder of this book, we will look at many other problems in the calculus of variations.

## 1.5. Recommended reading

Goldstine (1980), Fraser (2003), Kolmogorov and Yushkevich (1998), and Kline (1972) provide useful historical surveys of the calculus of variations.

Icaza Herrera (1994), Sussmann and Willems (1997), and Stein and Weichmann (2003) have written stimulating historical articles about the brachistochrone problem. An experimental study of the brachistochrone (using a "Hot Wheels" car) was carried out by Phelps et al. (1982).

The original 1697 solutions of John and Jacob Bernoulli can be found, translated into English, in Struik (1969). John Bernoulli's solution was recently reviewed by Erlichson (1999) and reviewed and generalized by Filobello-Nino et al. (2013).

If the endpoints $A$ and $B$ lie above the surface of the earth, but at vastly different heights, the gravitational field is no longer constant. One must instead determine the curve of swiftest descent in an attractive, inverse-square, gravitational field. This problem has been discovered repeatedly. Recent treatments include those of Singh and Kumar (1988), Parnovsky (1998), Tee (1999), and Hurtado (2000).

Goldstein and Bender (1986) analyzed the brachistochrone in the presence of relativistic effects and Farina (1987) showed that John Bernoulli's optical method can also be used to solve this relativistic problem. Kamath (1992) determined the relativistic tautochrone using fractional calculus.

The idea of high-speed tunnels through the earth is quite old. In Lewis Carroll's (1894) *Sylvie and Bruno Concluded*, Mein Herr describes a system of railway trains, without engines, powered by gravity:

> "Each railway is in a long tunnel, perfectly straight: so of course the *middle* of it is nearer the centre of the globe than the two ends: so every train runs half-way *down*-hill, and that gives it force enough to run the *other* half *up*-hill."

To which a protagonist replies:

> "Thank you. I understand that perfectly," said Lady Muriel. "But the velocity, in the *middle* of the tunnel, must be something *fearful!*"

You can also find a homework problem, about a tunnel-train between Minneapolis and Chicago, in Brooke and Wilcox (1929). See also Kirmser (1966).

Edwards' (1965) article reignited keen interest in gravity-powered transportation and inspired the articles by Cooper (1966a,b), Venezian (1966), Mallett (1966), Laslett (1966), and Patel (1967) on the terrestrial brachistochrone. Aravind (1981) applied John Bernoulli's optical method to the terrestrial brachistochrone and Prussing (1976), Chander (1977), McKinley (1979), and Denman (1985) pointed out that terrestrial brachistochrones are also tautochrones. Stalford and Garrett (1994) analyzed the terrestrial brachistochrone using differential geometry and optimal control theory.

Struik (1933), Carathéodory (1937), and Kline (1972) summarize the early history of the study of geodesics. Geodesics are an important topic in differential geometry (Struik, 1961; Oprea, 2007), Riemannian geometry (Berger, 2003), and geometric modeling (Patrikalakis and Maekawa, 2002). See Bliss (1902) for examples of geodesics on a toroidal anchor ring and Sneyd and Peskin (1990) for examples of geodesic trajectories on general tubular surfaces.

Isenberg (1992) and Oprea (2000) provide interesting and readable introductions to the science and mathematics of soap films. Barbosa and Colares (1986), Nitsche (1989), Fomenko (1990), and Fomenko and Tuzhilin (1991) do an excellent job of presenting the history and theory of minimal surfaces.

## 1.6. Exercises

**1.6.1. Descent time down a cycloidal curve.** Show that the descent time down the cycloidal curve

$$x(\phi) \;=\; a + R(\phi - \sin\,\phi)\,, \quad y(\phi) \;=\; y_a - R(1 - \cos\,\phi) \tag{1.48}$$

is

$$T \;=\; \sqrt{\frac{R}{g}}\,\phi_b\,, \tag{1.49}$$

where $\phi_b$ is the angle $\phi$ corresponding to the point $B = (b, y_b)$. What is the descent time to the lowest point on the cycloid?

**1.6.2. Complementary curves of descent.** The authors Mungan and Lipscombe (2013) recently introduced the term *complementary curves of descent* to describe curves that have identical descent times.

(a) Determine the descent time for a straight line (shown in bold in Figure 1.11).

(b) Rewrite integral (1.5) in polar coordinates assuming, for convenience, that $\theta$ increases clockwise.

(c) Determine the descent time for the lower portion of the leminiscate

$$r \;=\; 2\,c\sqrt{\sin\theta\,\cos\theta} \tag{1.50}$$

(shown in bold in Figure 1.11). Hint:

$$\frac{d}{d\theta}\left(\frac{\cos^{1/4}\theta}{\sin^{1/4}\theta}\right) \;=\; -\frac{1}{4}\,\cos^{-3/4}\theta\,\sin^{-5/4}\theta\,. \tag{1.51}$$

(d) Verify that the lemniscate is complementary to the straight line.

**1.6.3. Potential energy due to a spherical shell.** The gravitational potential energy between two point masses, $M$ and $m$, separated by a distance $r$ is

$$V(r) \;=\; -\frac{GMm}{r}\,, \tag{1.52}$$

where $G$ is the universal gravitational constant.

Calculate the potential energy of mass $m$ at point $P$ due to the gravitational attraction of a thin homogeneous spherical shell of mass

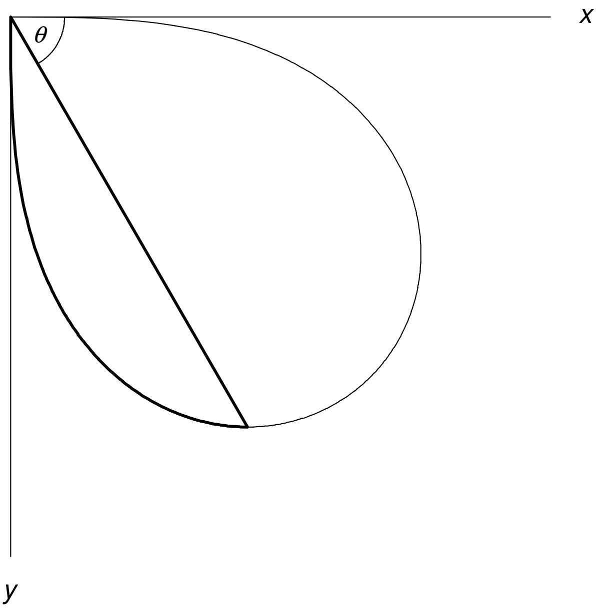



**Figure 1.11.** Complementary curves

$M$, surface (mass) density $\sigma$, and radius $x$ by integrating over a set of ring elements. (See Figure 1.12.) Assume that point $P$ is a distance $r$ from the center of the shell and that $y$ is the distance between the ring and point $P$. Be sure to consider the case when $P$ is inside the shell ($r < x$) as well as outside the shell ($r > x$).

**1.6.4. Potential energy inside the earth.** Use your results from the last problem and integrate over shells of appropriate radii to show that the potential energy of a point mass $m$ in a spherical and homogeneous earth can be written, to within an additive constant, as

$$V(r) = \frac{1}{2}\frac{mg}{R}\,r^2\,, \tag{1.53}$$

where $R$ is the radius of the earth, $g$ is the magnitude of the gravitational acceleration at the surface of the earth, $r$ is the distance of the point mass from the center of the earth, and $\rho$ is the (volumetric) density of the earth.

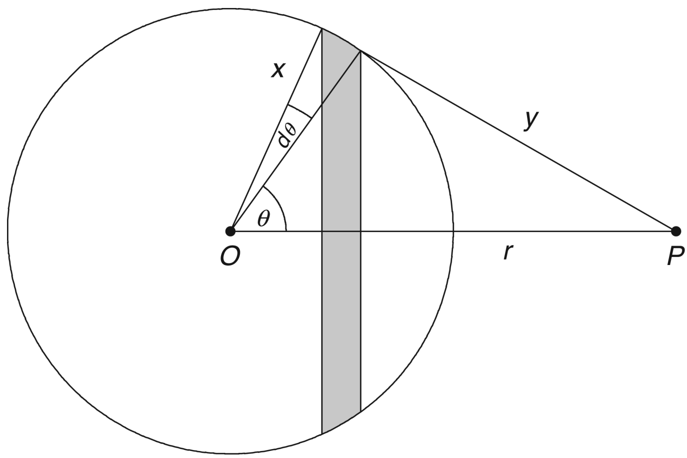



**Figure 1.12.** Geometry of a spherical shell

**1.6.5. Gauss's law.** Gauss's flux theorem for gravity states that the gravitational flux through a closed surface is proportional to the enclosed mass. Gauss's theorem can be written in differential form, using the divergence theorem, as

$$\nabla \cdot \mathbf{g} \;=\; -4\pi G\rho \,, \tag{1.54}$$

where $G$ is the universal gravitational constant, $\rho$ is the (volumetric) density of the enclosed mass, $\mathbf{g} = \mathbf{F}/m$ is the gravitational field intensity, $m$ is the mass of a test point, and $\mathbf{F}$ is the force on this test mass.

(a) Use this theorem to determine the force $\mathbf{F}(r)$ acting on mass $m$ at point $P$ due to the gravitational attraction of a thin homogeneous spherical shell of mass $M$, surface density $\sigma$, and radius $x$. Assume that point $P$ is a distance $r$ from the center of the shell. Be sure to consider the case where point $P$ is inside the shell ($r < x$) as well as outside the shell ($r > x$).

(b) Assume that $\mathbf{F}(r) = -dV/dr$, where $V(r)$ is the gravitational potential energy. Integrate the above force (starting at a reference point at infinity) to rederive the potential energy in Exercise 1.6.1.

(c) Use Gauss's flux theorem to determine the force $\mathbf{F}(r)$ acting on mass $m$ at point $P$ due to the gravitational attraction of a uniform solid sphere of mass $M$, density $\rho$, and radius $R$. Be sure to consider the case where point $P$ is inside the shell ($r < R$) as well as outside the shell ($r > R$).

(d) Integrate the above force (starting at a reference point at infinity) to rederive the potential energy in Exercise 1.6.2.

**1.6.6. First fundamental forms.** Determine the first fundamental form for *three* of the following seven surfaces. The surfaces you may choose from are:

(a) the helicoid

$$x = u\cos v\,, \quad y = u\sin v\,, \quad z = a\,v\,; \tag{1.55}$$

(b) the torus

$$x = (b + a\cos u)\cos v\,, \quad x = (b + a\cos u)\cos v\,, \quad z = a\sin u\,; \tag{1.56}$$

(c) the catenoid

$$x = a\cosh\frac{u}{a}\cos v\,, \quad y = a\cosh\frac{u}{a}\sin v\,, \quad z = u\,; \tag{1.57}$$

(d) the general surface of revolution

$$x = f(u)\cos v\,, \quad y = f(u)\sin v\,, \quad z = g(u)\,; \tag{1.58}$$

(e) the sphere (with alternate parameterization)

$$x = \frac{4a^2u}{4a^2 + u^2 + v^2}\,, \quad y = \frac{4a^2v}{4a^2 + u^2 + v^2}\,, \tag{1.59}$$

$$z = a\,\frac{4a^2 - u^2 - v^2}{4a^2 + u^2 + v^2}\,;$$

(f) the ellipsoid

$$x = a\cos u\cos v\,, \quad y = b\cos u\sin v\,, \quad z = c\sin u\,; \tag{1.60}$$

(g) the hyperbolic paraboloid

$$x = a(u + v)\,, \quad y = b(u - v)\,, \quad z = uv\,. \tag{1.61}$$

**1.6.7. Surface area.** Consider a surface written in the vector form

$$\mathbf{r}(u,v) \;=\; x(u,v)\,\mathbf{i} \;+\; y(u,v)\,\mathbf{j} \;+\; z(u,v)\,\mathbf{k}\,, \tag{1.62}$$

where $u$ and $v$ are parameters.

(a) Justify or motivate the surface-area formula

$$S \;=\; \iint ||\mathbf{r}_u \times \mathbf{r}_v||\; du\,dv\,. \tag{1.63}$$

(b) Show that the above surface-area formula can also be written as

$$S \;=\; \iint \sqrt{EG - F^2}\; du\,dv\,, \tag{1.64}$$

where $E$, $F$, and $G$ are the coefficients of the first fundamental form.

(c) Write the surface

$$z \;=\; f(x,y) \tag{1.65}$$

in vector form and show that the above formulas for area imply that

$$S \;=\; \iint\limits_{\Omega} \sqrt{1 + f_x^2 + f_y^2}\; dx\,dy\,. \tag{1.66}$$

**1.6.8. Surface area of a hyperbolic paraboloid.** Consider the hyperbolic paraboloid

$$\mathbf{r}(u,v) \;=\; u\,\mathbf{i} \;+\; v\,\mathbf{j} \;+\; uv\,\mathbf{k}\,. \tag{1.67}$$

Determine the surface area for that portion of the paraboloid that is specified by values of $u$ and $v$ that lie in the first quadrant of $(u,v)$ parameter space between the positive $u$- and $v$-axes and the circle

$$u^2 + v^2 \;=\; 1\,. \tag{1.68}$$

**1.6.9. Surface area of a helicoid.** Find the area of the portion of the helicoid

$$\mathbf{r}(u,v) \;=\; u\,\cos\,v\,\mathbf{i} \;+\; u\,\sin\,v\,\mathbf{j} \;+\; bv\,\mathbf{k} \tag{1.69}$$

that is specified by $0 \le u \le a$ and $0 \le v \le 2\pi$.

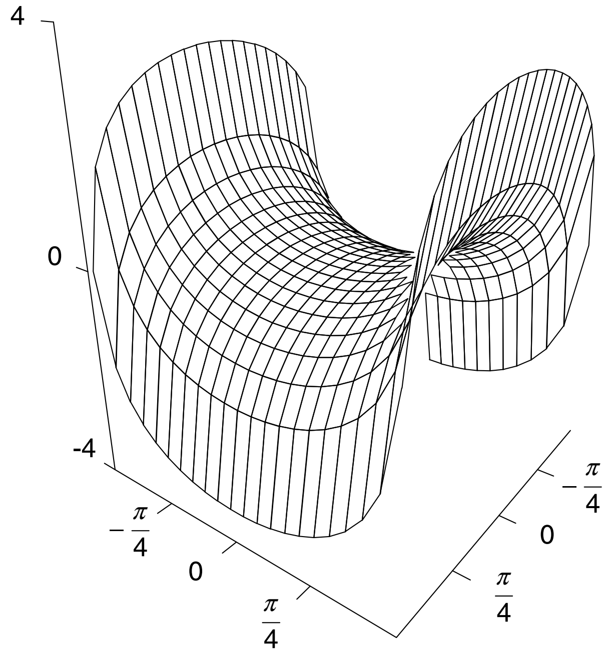



**Figure 1.13.** Scherk's first minimal surface

**1.6.10. Scherk's minimal surface.** Take the minimal surface equation, equation (1.47), and look for a solution of the form

$$f(x,y) \;=\; g(x) + h(y)\,. \tag{1.70}$$

Show that the resulting differential equation is separable. Solve for $g(x)$ and $h(y)$ to obtain *Scherk's (first) minimal surface,*

$$f(x,y) \;=\; c\,\ln\left[\frac{\cos\,(x/c)}{\cos\,(y/c)}\right]. \tag{1.71}$$

This surface was the first minimal surface discovered after the catenoid and the helicoid. A piece of this surface, for $c = 1$, $-\pi/2 < x < \pi/2$, and $-\pi/2 < y < \pi/2$, is shown in Figure 1.13.

Chapter 2

# The First Variation

## 2.1. The simplest problem

Our goal is to minimize (or to maximize) a definite integral of the form

$$J[y] \;=\; \int_a^b f(x, y(x), y'(x))\, dx \tag{2.1}$$

subject to the boundary conditions

$$y(a) \;=\; y_a\,, \quad y(b) \;=\; y_b\,. \tag{2.2}$$

I wrote $J[y]$ rather than $J(y)$ to emphasize that we are dealing with *functionals* and not just functions. Our definite integral returns a real number for each function $y(x)$. A functional is an operator that maps functions to real numbers. Functional analysis was, originally, the study of functionals. The purpose of the calculus of variations is to maximize or minimize functionals.

We will encounter functionals that act on all or part of several well-known function spaces. Function spaces that occur in the calculus of variations include the following:

(a) $C[a,b]$, the space of real-valued functions that are continuous on the closed interval $[a,b]$;

(b) $C^1[a, b]$, the space of real-valued functions that are continuous and that have continuous derivatives on the closed interval $[a, b]$;

(c) $C^2[a, b]$, the space of real-valued functions that are continuous and that have continuous first and second derivatives on the closed interval $[a, b]$;

(d) $D[a, b]$, the space of real-valued functions that are piecewise continuous on the closed interval $[a, b]$; and

(e) $D^1[a, b]$, the space of real-valued functions that are continuous and that have piecewise continuous derivatives on the closed interval $[a, b]$.

A piecewise continuous function can have a finite number of jump discontinuities in the interval $[a, b]$. The right-hand and left-hand limits of the function exist at the jump discontinuities. A function that is piecewise continuously differentiable is continuous but may have a finite number of corners.

We wish to find the *extremum* of a functional. Extremum is a word that was first introduced by Paul du Bois-Reymond (1879b). Du Bois-Reymond got tired of always having to say "maximum or minimum" and so he introduced a single term, extremum, to talk about both maxima and minima. The term stuck.

We will take our lead from (ordinary) calculus. We will look for a condition analogous to setting the first derivative equal to zero in calculus. The resulting *Euler–Lagrange equation* is quite important, so much so that we will derive this equation in three ways. We will begin with Euler's heuristic derivation (Euler, 1744) and then move on to Lagrange's 1755 derivation (the traditional approach). We will then consider Paul du Bois-Reymond's modification of Lagrange's derivation (du Bois-Reymond, 1879a).

## 2.2. Euler's approach

Leonhard Euler was the first person to systematize the study of variational problems. His 1744 opus, *A Method for Finding Curved Lines Enjoying Properties of Maximum or Minimum, or Solution of Isoperimetric Problems in the Broadest Accepted Sense*, is a compendium of 100 special problems. The book also contains a general method for

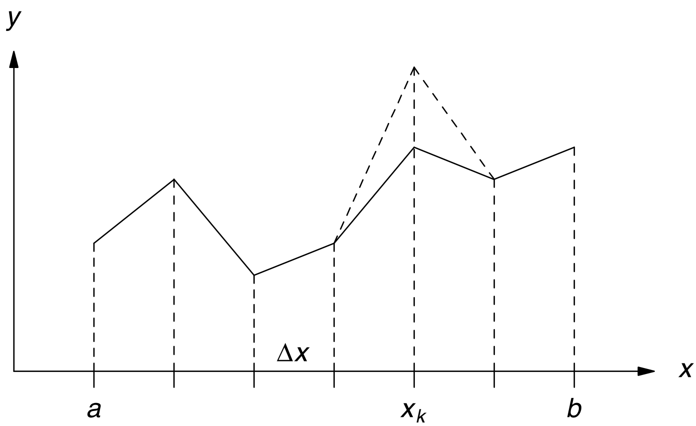



**Figure 2.1.** Polygonal curves

handling these problems. Euler dropped his method for Lagrange's more elegant "method of variations" after receiving Lagrange's (August 12, 1755) letter. Euler also named this subject the *calculus of variations* in Lagrange's honor.

Euler's essential idea was to first go from a variational problem to an $n$-dimensional problem and to then pass to the limit as $n \to \infty$. We will borrow from the modernized treatment of Euler's method found in Elsgolc (1961) and Gelfand and Fomin (1963). See Goldstine (1980) and Fraser (2003) for more on the original approach.

Let us divide the closed interval $[a, b]$ into $n+1$ equal subintervals (see Figure 2.1). We will assume that the subintervals are bounded by the points

$$x_0 = a, x_1, \ldots, x_n, x_{n+1} = b\,. \tag{2.3}$$

Each subinterval is of width

$$\Delta x = x_{i+1} - x_i = \frac{(b-a)}{n+1}\,. \tag{2.4}$$

We will also replace the smooth function $y(x)$ by the polygonal curve with vertices

$$(x_0, y_0),\ (x_1, y_1), \ldots,\ (x_n, y_n),\ (x_{n+1}, y_{n+1})\,. \tag{2.5}$$

Here, $y_i = y(x_i)$. We can now approximate the functional $J[y]$ by the sum

$$J(y_1, \ldots, y_n) \equiv \sum_{i=0}^{n} f\left(x_i, y_i, \frac{y_{i+1} - y_i}{\Delta x}\right) \Delta x\,, \qquad (2.6)$$

a function of $n$ variables. (Remember that $y_0 = y_a$ and $y_{n+1} = y_b$ are fixed.)

What is the effect of raising or lowering one of the free $y_i$? To answer this question, let us choose one of the free $y_i$, $y_k$, and take the partial derivative with respect to $y_k$. Since $y_k$ appears in only two terms in our sum, the partial derivative is just

$$\begin{aligned}\frac{\partial J}{\partial y_k} &= f_y\left(x_k, y_k, \frac{y_{k+1} - y_k}{\Delta x}\right) \Delta x \qquad (2.7)\\ &\quad + f_{y'}\left(x_{k-1}, y_{k-1}, \frac{y_k - y_{k-1}}{\Delta x}\right)\\ &\quad - f_{y'}\left(x_k, y_k, \frac{y_{k+1} - y_k}{\Delta x}\right).\end{aligned}$$

To find an extremum, we would ordinarily set this partial derivative equal to zero for each $k$. We also, however, want to take the limit as $n \to \infty$. In this limit, $\Delta x \to 0$ and the right-hand side of equation (2.7) goes to zero. The equation $0 = 0$, while true, is, sadly, not very helpful. To obtain a nontrivial result, we must first divide by $\Delta x$,

$$\begin{aligned}\frac{1}{\Delta x}\frac{\partial J}{\partial y_k} &= f_y\left(x_k, y_k, \frac{y_{k+1} - y_k}{\Delta x}\right) \qquad (2.8)\\ -\frac{1}{\Delta x}&\left[f_{y'}\left(x_k, y_k, \frac{y_{k+1} - y_k}{\Delta x}\right) - f_{y'}\left(x_{k-1},\, y_{k-1}, \frac{y_k - y_{k-1}}{\Delta x}\right)\right].\end{aligned}$$

As we now let $n \to \infty$ and $\Delta x \to 0$, equation (2.8) yields the *variational derivative*

$$\frac{\delta J}{\delta y} = f_y(x, y, y') - \frac{d}{dx} f_{y'}(x, y, y')\,. \qquad (2.9)$$

This variational derivative plays the same role for functionals that the partial derivative plays for multivariate functions. For a relative (or local) minimum, we expect this derivative to vanish at each point, leaving us with the *Euler–Lagrange equation*

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) \;=\; 0\,. \tag{2.10}$$

This condition must be modified if the minimizing curve lies on the boundary rather than in the interior of the region of interest. Moreover, the Euler–Lagrange equation is only a *necessary* condition, in the same sense that $f'(x) = 0$ is a necessary, but not a sufficient, condition in calculus.

I should, perhaps, add that the above discussion is misleading to the extent that the formal notion of a variational or functional derivative was not introduced until much later, by Vito Volterra (1887), in the early stages of the development of functional analysis. See the recommended reading at the end of this chapter for more information about variational derivatives.

**Example 2.1** (Shortest curve in the plane)**.**

Let's see what the Euler–Lagrange equation has to say about the shape of the shortest curve between two points, $(a, y_a)$ and $(b, y_b)$, in the plane. We clearly wish to minimize the arc-length functional

$$J[y] \;=\; \int_a^b ds \;=\; \int_a^b \sqrt{1+y'^2}\; dx\,. \tag{2.11}$$

The integrand,

$$f(x, y, y') \;=\; \sqrt{1+y'^2}\,, \tag{2.12}$$

does not depend on $y$ and so the Euler–Lagrange equation reduces to

$$\frac{d}{dx}\left(\frac{y'}{\sqrt{1+y'^2}}\right) \;=\; 0\,. \tag{2.13}$$

Integrating once produces

$$\frac{y'}{\sqrt{1+y'^2}} \;=\; \text{constant} \tag{2.14}$$

and we quickly conclude that

$$y' \;=\; c\,, \tag{2.15}$$

$c$ a constant. If we integrate once again and set our new constant of integration to $d$, we conclude that

$$y = cx + d\,. \tag{2.16}$$

This is the equation of a straight line. The constants $c$ and $d$ can be determined from the boundary conditions.

## 2.3. Lagrange's approach

Let us return to the problem of minimizing or maximizing the functional

$$J[y] = \int_a^b f(x, y(x), y'(x))\, dx \tag{2.17}$$

subject to the boundary conditions

$$y(a) = y_a\,, \quad y(b) = y_b\,. \tag{2.18}$$

Euler derived the Euler–Lagrange equation by varying a single ordinate. Lagrange realized that he could derive this same equation while simultaneously varying *all* of the (free) ordinates.

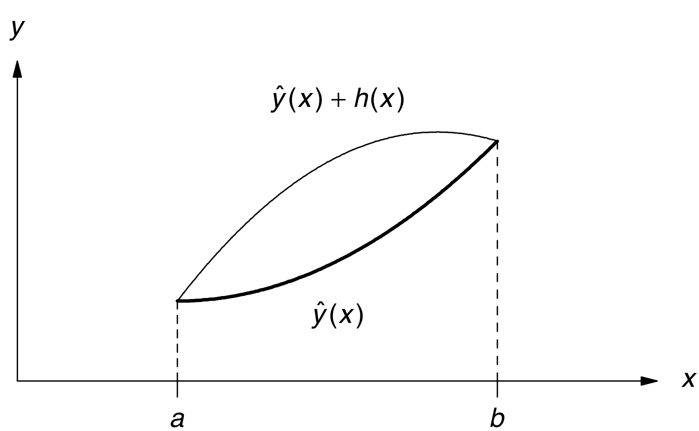



**Figure 2.2.** A small variation

Let us suppose that the function $y = \hat{y}(x)$ solves our problem. We now introduce $h(x)$, a small deviation or *variation* from this idealized solution,

$$y(x) = \hat{y}(x) + h(x) \tag{2.19}$$

(see Figure 2.2), that satisfies

$$h(a) = 0 \text{ and } h(b) = 0\,. \tag{2.20}$$

At this point, we need to discuss a subtle point that escaped Lagrange but that turns out to be rather important. What exactly do we mean when we say that a variation is small? The usual way to measure the nearness of two functions is to compute the *norm* of the difference of the two functions. There are many possible norms and we will see that our conclusions about extrema (maxima and minima) are rather sensitive to which norm we use.

We will use two different norms throughout this course. They are the *weak* norm

$$||h||_w = \max_{[a,b]} |h(x)| \tag{2.21}$$

and the *strong* norm

$$||h||_s = \max_{[a,\,b]} |h(x)| + \sup_{[a,b]} |h'(x)|\,. \tag{2.22}$$

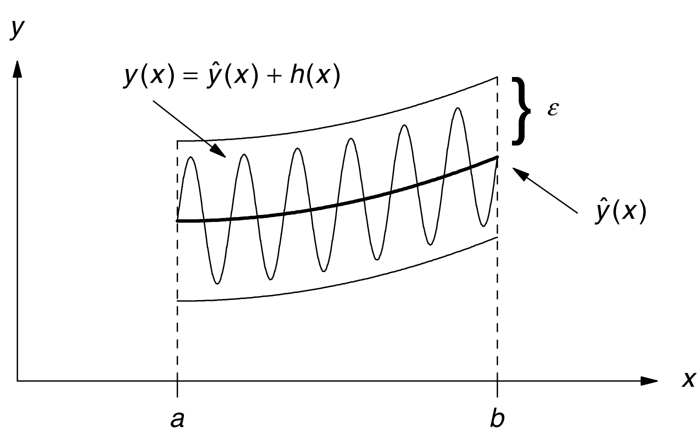



**Figure 2.3.** Strong variation

The supremum (or least upper bound) is here in case we are working with functions that are piecewise continuously differentiable. If our functions are continuously differentiable, the supremum can be replaced by a maximum.

We will use the weak and strong norms to establish neighborhoods in function space. Weak and strong norms permit different variations about the optimal solution. Since the weak norm does not impose any restriction upon the derivative, an $\epsilon$-neighborhood in a weakly-normed space will include *strong variations* (see Figure 2.3) that differ significantly from the optimal solution in slope while remaining close in ordinate.

Strong variations may have arbitrarily large derivatives.

**Example 2.2.**

The function

$$h(x) \;=\; \epsilon \sin\left(\frac{x}{\epsilon^2}\right) \tag{2.23}$$

never exceeds $\epsilon$ and yet its derivative,

$$h'(x) \;=\; \frac{1}{\epsilon} \cos\left(\frac{x}{\epsilon^2}\right), \tag{2.24}$$

may become arbitrarily large as $\epsilon$ is made small.

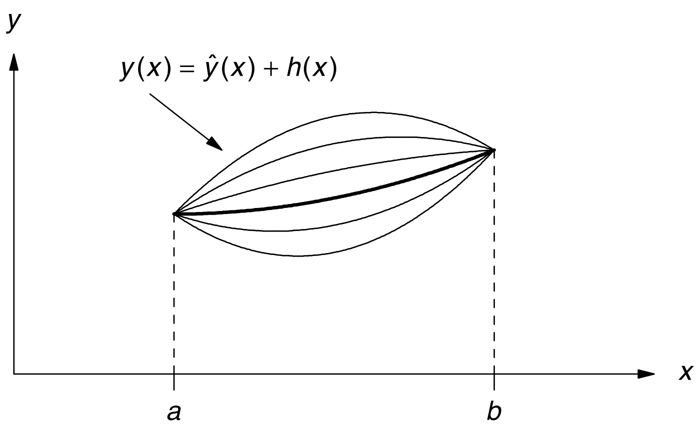



**Figure 2.4.** Weak variations

The strong norm, in contrast, *does* place a restriction on the size of the derivative. Stating that

$$||h||_s \ < \ \epsilon \tag{2.25}$$

implies not only

$$\max_{[a,b]} \ |h(x)| \ < \ \epsilon \tag{2.26}$$

but also

$$\sup_{[a,b]} \ |h'(x)| \ < \ \epsilon \, . \tag{2.27}$$

An $\epsilon$-neighborhood in a strongly-normed space contains only weak variations (see Figure 2.4) that are close to the optimal solution in both ordinate *and* slope.

Since strong variations are a superset of weak variations, a function that minimizes a functional relative to nearby strong variations also minimizes that functional relative to nearby weak variations. Conversely, a necessary condition for a weak relative minimum is also a necessary condition for a strong relative minimum. Lagrange's approach uses weak variations. This is alright if we want necessary conditions but is a problem if we want sufficient conditions. In due course, we will encounter examples of functionals that have minima relative to weak variations, but not relative to strong variations.

To make Lagrange's assumption as explicit as possible, we will consider small weak variations

$$h(x) \ = \ \epsilon \, \eta(x) \tag{2.28}$$

where

$$\eta(a) \ = \ 0 \, , \quad \eta(b) \ = \ 0 \tag{2.29}$$

and $h(x)$ and $h'(x)$ are of the same order of smallness. The function $\eta(x)$ is thus assumed to be independent of the parameter $\epsilon$. As $\epsilon$ tends to zero, the variation $h(x)$ tends to zero in both ordinate and slope. For notational convenience, we will also think of the functional $J[y]$ as a function of $\epsilon$,

$$J(\epsilon) \ \equiv \ J[\hat{y} + \epsilon\eta] \ = \ \int_a^b f(x, \hat{y} + \epsilon\eta, \hat{y}' + \epsilon\eta') \ dx \, . \tag{2.30}$$

Let us now look at the *total variation*

$$\Delta J \;=\; J(\epsilon) - J(0)\,. \tag{2.31}$$

That is,

$$\begin{aligned}\Delta J \;&=\; \int_a^b f(x, \hat{y} + \epsilon\eta,\, \hat{y}' + \epsilon\eta')\; dx - \int_a^b f(x, \hat{y}, \hat{y}')\; dx \qquad (2.32)\\ &=\; \int_a^b [f(x, \hat{y} + \epsilon\eta, \hat{y}' + \epsilon\eta') - f(x, \hat{y}, \hat{y}')]\; dx\,.\end{aligned}$$

If $f$ has enough continuous partial derivatives — and we shall assume that it does — we may expand the total variation in a power series in $\epsilon$. Using the usual Taylor expansion, we obtain

$$\Delta J \;=\; \delta J \,+\, \frac{1}{2}\,\delta^2 J \,+\, O(\epsilon^3)\,. \tag{2.33}$$

Here,

$$\begin{aligned}\delta J \;&=\; \left.\frac{dJ(\epsilon)}{d\epsilon}\right|_{\epsilon\,=\,0} \epsilon \qquad (2.34)\\ &= \epsilon \int_a^b [f_y(x, \hat{y}, \hat{y}')\,\eta + f_{y'}\,(x, \hat{y}, \hat{y}')\,\eta']\; dx\end{aligned}$$

is the *first variation.* Likewise,

$$\begin{aligned}\delta^2 J \;&=\; \left.\frac{d^2 J(\epsilon)}{d\epsilon^2}\right|_{\epsilon\,=\,0} \epsilon^2 \qquad (2.35)\\ &= \epsilon^2 \int_a^b [f_{yy}(x, \hat{y}, \hat{y}')\,\eta^2 + 2\, f_{yy'}\,(x, \hat{y}, \hat{y}')\,\eta\,\eta' + f_{y'y'}\,(x, \hat{y}, \hat{y}')\,{\eta'}^2]\; dx\end{aligned}$$

is the *second variation.* For $\epsilon$ sufficiently small, we expect that a nonvanishing first variation will dominate the right-hand side of total variation (2.33). Likewise, we expect that a nonvanishing second variation will dominate higher-order terms.

If $J[\hat{y}]$ is a relative (or local) minimum, we must have

$$\Delta J \;\geq\; 0 \tag{2.36}$$

for all sufficiently small $\epsilon$. Since, however, the first variation is odd in $\epsilon$, we can change its sign by changing the sign of $\epsilon$. To prevent this change in sign, we require that

$$\delta J \;=\; 0\,. \tag{2.37}$$

For a minimum, we also require that

$$\delta^2 J \;\geq\; 0\,. \tag{2.38}$$

If we want a relative maximum, we will, in turn, require

$$\delta J \,=\, 0\,, \quad \delta^2 J \,\leq\, 0\,. \tag{2.39}$$

It is convenient, at this early stage of the course, to focus on the first variation. In light of the above arguments, we may safely say:

> ***First variation condition:***
> A necessary condition for the functional $J[y]$ to have a relative (or local) minimum or maximum at $y = \hat{y}(x)$ is that the first variation of $J[y]$ vanish,
> $$\delta J \;=\; 0\,, \tag{2.40}$$
> for $y = \hat{y}(x)$ and for all admissible variations $\eta(x)$.

The first variation,

$$\delta J \;=\; \epsilon \int_a^b \left[ f_y(x,\hat{y},\hat{y}')\,\eta + f_{y'}\,(x,\hat{y},\hat{y}')\,\eta' \right]\,dx\,, \tag{2.41}$$

is rather unwieldy as written. We will rewrite the first variation so as to factor out the dependence on the admissible variations $\eta(x)$. There are two different ways to do this. Both methods involve integration by parts. We start with Lagrange's approach.

**2.3.1. Lagrange's simplification.** Let us subject the second term in integrand (2.41) to integration by parts,

$$\int_a^b f_{y'}\,(x,\hat{y},\hat{y}')\eta'\;dx \;=\; \eta(x)\left.\frac{\partial f}{\partial y'}\right|_{x\,=\,a}^{x\,=\,b} - \int_a^b \eta\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\,dx\,. \tag{2.42}$$

Since our variations from the idealized solution vanish at the endpoints of the interval,

$$\eta(a) = 0\,, \quad \eta(b) = 0\,, \tag{2.43}$$

our first necessary condition reduces to

$$\epsilon \int_a^b \eta(x) \left[ \frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right]_{\hat{y},\hat{y}'} dx = 0 \tag{2.44}$$

for all admissible $\eta(x)$. The subscript in this last equation signifies that the expression in square brackets is evaluated at $y = \hat{y}(x)$ and $y' = \hat{y}'(x)$.

Let us note, right away, that our use of integration by parts, in this way, pretty much forces us to assume that $\hat{y}(x)$ is twice differentiable. The partial derivative $f_{y'}$ is generally a function of $y'$ (as well as of $y$ and $x$) and if $y''$ does not exist, the existence of

$$\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) \tag{2.45}$$

becomes doubtful. We shall see, momentarily, that Lagrange's simplification actually forces us to assume that $\hat{y}''(x) \in C[a,b]$ or $\hat{y}(x) \in C^2\,[a,b]$.

Lagrange claimed, without proof, that the coefficient of $\eta(x)$ in equation (2.44) must vanish, yielding the Euler–Lagrange equation,

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0\,. \tag{2.46}$$

Euler pointed out, in a communication to Lagrange, that Lagrange's statement was not self-evident and that he really ought to *prove* that the coefficient of $\eta(x)$ must vanish. This proof was eventually supplied by du Bois-Reymond (1879a). Du Bois-Reymond's result is now known as the fundamental lemma of the calculus of variations.

***Fundamental lemma of the calculus of variations:***
If $M(x) \in C[a,b]$ and if

$$\int_a^b M(x)\,\eta(x)\;dx \;=\; 0 \tag{2.47}$$

for every $\eta(x) \in C^1\,[a,b]$ such that

$$\eta(a) \;=\; \eta(b) \;=\; 0\,, \tag{2.48}$$

then

$$M(x) \;=\; 0 \tag{2.49}$$

for all $x \in [a,b]$.

**Proof.** The proof is by contradiction. Suppose (without loss of generality) that $M(x)$ is positive at some point in $(a,b)$. $M(x)$ must then, by continuity, be positive in some interval $[x_1,x_2]$ within $[a,b]$. Now (see Figure 2.5), let

$$\eta(x) \;=\; \begin{cases} (x-x_1)^2\,(x-x_2)^2\,, & x \in [x_1,x_2]\,, \\ 0\,, & x \notin [x_1,x_2]\,. \end{cases} \tag{2.50}$$

Clearly, $\eta(x) \in C^1[a,b]$. With this choice of $\eta(x)$,

$$\int_a^b M(x)\,\eta(x)\,dx \;=\; \int_{x_1}^{x_2} M(x)(x-x_1)^2(x-x_2)^2\;dx\,. \tag{2.51}$$

Since the integrand is nonnegative,

$$\int_a^b M(x)\,\eta(x)\;dx \;>\; 0\,. \tag{2.52}$$

This in contrary to our original hypothesis and it now follows that

$$M(x) \;=\; 0\,, \quad x \in (a,b)\,. \tag{2.53}$$

The continuity of $M(x)$, in turn, guarantees that $M(x)$ also vanishes at the endpoints of the interval. ♣

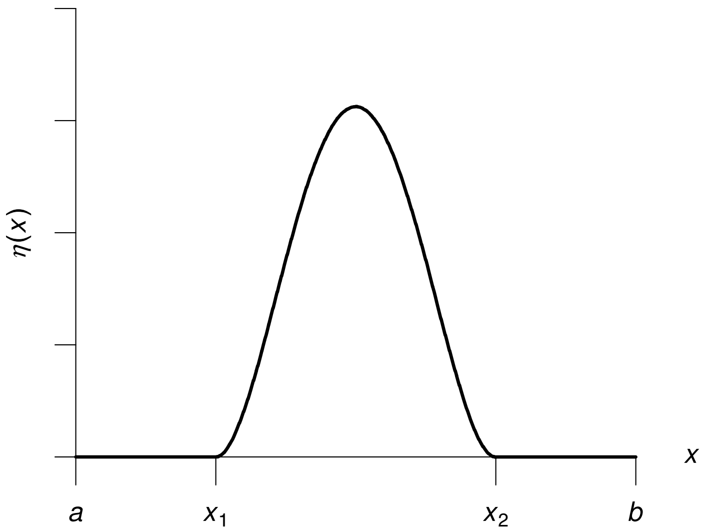



**Figure 2.5.** A nonnegative bump

To be able to apply this fundamental lemma of the calculus of variations, we must be sure that

$$M(x) \;=\; \frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) \tag{2.54}$$

is continuous on the closed interval $[a,\, b]$. If we apply the chain rule, we may rewrite the right-hand side of this last equation in the *ultra-differentiated* form

$$\begin{aligned} M(x) \;&=\; \frac{\partial f}{\partial y} - \frac{\partial f_{y'}}{\partial x}\,\frac{dx}{dx} - \frac{\partial f_{y'}}{\partial y}\,\frac{dy}{dx} - \frac{\partial f_{y'}}{\partial y'}\,\frac{dy'}{dx} \\ &=\; f_y - f_{y'x} - f_{y'y}\,y' - f_{y'y'}\,y''\,. \end{aligned} \tag{2.55}$$

To obtain the Euler–Lagrange equation using Lagrange's simplification, we must therefore make the additional assumption that $\hat{y}''(x) \in C[a,b]$ or that $\hat{y}(x) \in C^2[a,b]$.

Having made (or, more honestly, having been forced into) the assumption that $\hat{y}(x) \in C^2\,[a,b]$, we can now state the following *necessary* condition for a relative maximum or minimum:

***Euler–Lagrange condition:***
Every $\hat{y}(x) \in C^2[a,b]$ that produces a relative extremum of the integral

$$J[y] \;=\; \int_a^b f(x,y,y')\,dx \tag{2.56}$$

satisfies the Euler–Lagrange differential equation

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) \;=\; 0\,. \tag{2.57}$$

Lagrange's simplification forces us to *assume* that our solutions have continuous second derivatives. Can we loosen this assumption? Let us start with the necessary condition that the first variation must vanish,

$$\delta J[\eta] \;=\; \epsilon \int_a^b \left[f_y(x,\hat{y},\hat{y}')\,\eta + f_{y'}(x,\hat{y},\hat{y}')\,\eta'\right]\,dx \;=\; 0\,, \tag{2.58}$$

and try a different approach.

**2.3.2. Du Bois-Reymond's simplification.** Let us now assume that the functions $\hat{y}(x)$ and $\eta(x)$ are merely continuously differentiable, $\hat{y}(x)$, $\eta(x) \in C^1[a,b]$. Since $f_{y'}(x,\hat{y},\hat{y}')$ depends on $\hat{y}'(x)$, this function need not be differentiable. As a result, we cannot integrate the second term in integrand (2.41) by parts.

Let us instead integrate the first term in integrand (2.41) by parts. Doing so, we obtain

$$\int_a^b f_y(x,\hat{y},\hat{y}')\,\eta\,dx \;=\; \left[\eta(x)\phi(x)\right]_{x\,=\,a}^{x\,=\,b} - \int_a^b \phi(x)\,\eta'(x)\,dx\,, \tag{2.59}$$

where

$$\phi(x) \;=\; \int_a^x f_y(u,\hat{y}(u),\hat{y}'(u))\,du\,. \tag{2.60}$$

Since we have only assumed the continuity of $f_y(x, \hat{y}, \hat{y}')$ and of $\eta'(x)$, this integration by parts *is* legal. Since

$$\eta(a) \;=\; \eta(b) \;=\; 0\,, \tag{2.61}$$

necessary condition (2.58) now reduces to

$$\int_a^b \left( \frac{\partial f}{\partial y'} - \int_a^x \frac{\partial f}{\partial y}\, du \right)_{\hat{y},\hat{y}'} \eta'(x)\; dx \;=\; 0\,. \tag{2.62}$$

We clearly need another lemma to progress further. Here it is:

***Lemma of du Bois-Reymond:***
If $M(x) \in C[a, b]$ and

$$\int_a^b M(x)\,\eta'(x)\; dx \;=\; 0 \tag{2.63}$$

for all $\eta(x) \in C^1[a, b]$ such that

$$\eta(a) \;=\; \eta(b) \;=\; 0\,, \tag{2.64}$$

then

$$M(x) \;=\; c\,, \tag{2.65}$$

a constant, for all $x \in [a, b]$.

**Proof.** We may prove this lemma by considering one well-chosen variation $\eta(x)$. Let $\mu$ denote the mean value of $M(x)$ on the closed interval $[a, b]$,

$$\mu \;=\; \frac{1}{(b-a)} \int_a^b M(x)\; dx\,. \tag{2.66}$$

Clearly,

$$\int_a^b [M(x) - \mu]\; dx \;=\; 0\,. \tag{2.67}$$

Now, consider the variation $\eta(x)$ defined by the equation

$$\eta(x) \;=\; \int_a^x [M(u) - \mu] \; du\,. \tag{2.68}$$

It is easy to see that $\eta(x) \in C^1[a,\, b]$. The function $\eta(x)$ also vanishes at $x = a$ and $x = b$. It is clearly an admissible variation. Moreover,

$$\eta'(x) \;=\; M(x) - \mu\,. \tag{2.69}$$

By hypothesis,

$$\int_a^b M(x)\,\eta'(x)\; dx \;=\; \int_a^b M(x)\,[M(x) - \mu]\; dx \;=\; 0\,. \tag{2.70}$$

Also,

$$\int_a^b M(x)\,[M(x) - \mu]\; dx - \mu \int_a^b [M(x) - \mu]\; dx \;=\; 0\,. \tag{2.71}$$

But, this last equation may be rewritten

$$\int_a^b [M(x) - \mu]^2\; dx \;=\; 0\,. \tag{2.72}$$

Let $x_0 \in [a, b]$ be a point where $M(x)$ is continuous. If $M(x_0) \neq \mu$, then there would have to exist a subinterval about $x = x_0$ on which $M(x) \neq \mu$. But this is clearly impossible in light of our last displayed equation. Thus $M(x) = \mu$ at all points of continuity. It follows that $M(x)$ is constant for all $x \in [a, b]$. ♣

We now wish to apply this lemma to necessary condition (2.62),

$$\int_a^b \left( \frac{\partial f}{\partial y'} - \int_a^x \frac{\partial f}{\partial y}\; du \right)_{\hat{y},\hat{y}'} \eta'(x)\; dx \;=\; 0\,. \tag{2.73}$$

Note that

$$M(x) \;=\; \frac{\partial f}{\partial y'} - \int_a^x \frac{\partial f}{\partial y}\, du \tag{2.74}$$

is continuous on $[a, b]$ and that the assumptions of the lemma are satisfied. It now follows that

$$\frac{\partial f}{\partial y'} \;=\; \int_a^x \frac{\partial f}{\partial y}\, du + c \tag{2.75}$$

for all $x \in [a, b]$. This is known as the *integrated form* of the Euler–Lagrange equation.

The right-hand side of equation (2.75) is differentiable. This, in turn, implies that the left-hand side of equation (2.75) is differentiable and that $\hat{y}(x)$ satisfies the Euler–Lagrange equation,

$$\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) \;=\; \frac{\partial f}{\partial y}\,. \tag{2.76}$$

In other words, all solutions with continuous first derivatives, not just those with continuous second derivatives, must satisfy the Euler–Lagrange equation.

The differentiability of $f_{y'}(x, \hat{y}, \hat{y}')$ can also be used to *prove* (Whittemore, 1900–1901) the existence of the second derivative $\hat{y}''(x)$ for all values of $x$ for which $f_{y'y'}(x, \hat{y}, \hat{y}') \neq 0$.

We will see later that we can weaken the conditions on $\hat{y}(x)$ and $\eta(x)$ even further, so that they are merely piecewise continuously differentiable. One can then show that the Euler–Lagrange equation is satisfied between corners of the solution and that additional conditions must be satisfied at the corners. Determining these conditions requires additional tools, and so we will defer the topic of corners until Chapter 10. For the time being, we will focus our attention on continuously differentiable solutions.

## 2.4. Recommended reading

Goldstine (1980), Fraser (1994, 2005a), and Thiele (2007) analyze Euler's early contributions to the calculus of variations. Euler's idea of using a polygonal curve to approximate the solution of a variational problem was revived in the 20th century by Russian mathematicians working on *direct methods* of solution. In a direct method, you construct a sequence of approximating functions, determine the unknown values and coefficients in each function using minimization, and let the sequence of functions converge to the solution. Euler's approach suggests the *direct method of finite differences* (Elsgolc, 1961). Other direct methods include the Ritz method, which is frequently and inappropriately (Leissa, 2005) called the Rayleigh-Ritz method, the Kantorovich method, and the Galerkin method. See Forray (1968) for an introduction to these direct methods.

The theory of the differentiation of functionals has its origins in the work of Volterra (1887). See Gelfand and Fomin (1963), Hamilton and Nashed (1982, 1995), Kolmogorov and Yushkevich (1998), and Lebedev and Cloud (2003) for more on variational derivatives. There are errors in the statements and proofs of the existence of the variational derivative for the simplest problem of the calculus of variations in Volterra (1913) and Gelfand and Fomin (1963). These errors were pointed out and corrected by Bliss (1915) and Hamilton (1980).

Lagrange announced his new approach to the calculus of variations in a 1755 letter to Euler; his results appeared in print seven years later (Lagrange, 1762). Fraser (1985) reviews the lengthy correspondence between Joseph Lagrange and Leonhard Euler and traces the development of Lagrange's approach to the foundations of the calculus of variations.

It has been said that the fundamental lemma is like a watchdog that guards the entrance gates to the entire classical domain of the calculus of variations (Dresden, 1932). Many early writers took the conclusion of the fundamental lemma as self-evident while others erred in their proof of this lemma. Huke (1931) traces the long and fascinating history of the fundamental lemma and of the lemma of du Bois-Reymond.

In writing this chapter, we leaned heavily on Bolza (1973) and Sagan (1969). I encourage all students of the calculus of variations to read these two books.

## 2.5. Exercises

**2.5.1. Euler's approach.** Using Euler's approach from Section 2.2, determine polygonal approximations to the curve that minimizes

$$\int_0^2 [(y')^2 + 6x^2 y]\, dx \tag{2.77}$$

subject to

$$y(0) = 2, \quad y(2) = 4 \tag{2.78}$$

for $n = 1$, $n = 2$, and $n = 3$. Write down and solve the Euler–Lagrange equation for this problem. Compare your polygonal approximations to your solution of the Euler–Lagrange equation.

**2.5.2. Another lemma.** Let $M(x) \in C[a, b]$ be a continuous function on the closed interval $a \leq x \leq b$ that satisfies

$$\int_a^b M(x)\, \eta''(x)\, dx = 0 \tag{2.79}$$

for all $\eta(x) \in C^2[a, b]$ satisfying

$$\eta(a) = \eta(b) = \eta'(a) = \eta'(b) = 0\,. \tag{2.80}$$

Prove that

$$M(x) = c_0 + c_1 x \tag{2.81}$$

for suitable constants $c_0$ and $c_1$. What can you say about $c_0$ and $c_1$?

Chapter 3

# Cases and Examples

## 3.1. Special cases

We are interested in the Euler–Lagrange equation,

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) \;=\; 0\,, \tag{3.1}$$

and in the curves that satisfy this equation. We will refer to these curves as *extremals* or, more rarely, as *stationary points* of the functional. The term extremal was introduced by Adolph Kneser (1900). Do not confuse *extremals,* solutions of the Euler–Lagrange equation, with *extrema*, the term introduced by du Bois-Reymond (1879b) for maxima or minima.

To write out the Euler–Lagrange equation in its explicit form, we must take into account the fact that $f_{y'}$ is a function of three variables, $x$, $y$, and $y'$, and that $y$ and $y'$ are functions of $x$. Therefore

$$\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) \;=\; \frac{\partial f_{y'}}{\partial x}\,\frac{dx}{dx} + \frac{\partial f_{y'}}{\partial y}\,\frac{dy}{dx} + \frac{\partial f_{y'}}{\partial y'}\,\frac{dy'}{dx} \tag{3.2}$$

so that the Euler–Lagrange equation may be written in the *ultra-differentiated* form

$$f_y - f_{y'x} - f_{y'y}y' - f_{y'y'}y'' \;=\; 0\,. \tag{3.3}$$

This means that the Euler–Lagrange equation is, generally speaking, a second-order ordinary differential equation. Since

$$y'' = \frac{1}{f_{y'y'}} \left(f_y - f_{y'x} - f_{y'y}y'\right), \tag{3.4}$$

a great deal will depend on whether $f_{y'y'}$ vanishes or not. If $f_{y'y'}$ never vanishes, we have a *regular problem.* If $f_{y'y'}$ always vanishes, we have the first of our special cases.

**3.1.1. Degenerate functionals.** This case occurs if the integrand of our functional is either independent of $y'$,

$$f(x, y, y') = M(x, y), \tag{3.5}$$

or depends linearly on $y'$,

$$f(x, y, y') = M(x, y) + N(x, y)\, y'. \tag{3.6}$$

(The former is a special case of the latter.) Functionals for which this is true are referred to as *degenerate* functionals. The Euler–Lagrange equation for a degenerate functional takes the form

$$\frac{\partial M}{\partial y} + \frac{\partial N}{\partial y}\, y' - \frac{\partial N}{\partial x} - \frac{\partial N}{\partial y}\, y' = 0 \tag{3.7}$$

or

$$\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x} = 0. \tag{3.8}$$

This equation does not contain derivatives of the unknown function and is not a differential equation.

Several subcases arise in practice. In the first subcase, the Euler–Lagrange equation (3.8) gives rise to one or more curves. One or another of these curves may satisfy the boundary conditions. More commonly, however, the curves fail to satisfy one or both of the boundary conditions and we cannot solve our problem, as specified.

**Example 3.1.**

The functional

$$J[y] = \int_0^1 \left[\frac{1}{2}(x - y)^2 + (\sin y)\, y'\right] dx \tag{3.9}$$

has the Euler–Lagrange equation

$$y - x \;=\; 0\,. \tag{3.10}$$

The solution to this algebraic equation,

$$y \;=\; x\,, \tag{3.11}$$

fails to satisfy most boundary conditions.

**Example 3.2.**

The functional

$$J[y] \;=\; \int_{-1}^{1} \cos y \; dx \tag{3.12}$$

has the Euler–Lagrange equation

$$-\sin y \;=\; 0\,. \tag{3.13}$$

An extremum may occur along one of the horizontal lines

$$y \;=\; n\,\pi\,, \quad n \;=\; 0,\, \pm 1,\, \pm 2, \ldots\,, \tag{3.14}$$

but only if the horizontal line satisfies the boundary conditions.

More rarely, the equation

$$\frac{\partial M}{\partial y} \;=\; \frac{\partial N}{\partial x} \tag{3.15}$$

is satisfied identically for any $y(x)$ and the functional is independent of the path of integration. Indeed, for the integrand

$$f(x, y, y') \;=\; M(x, y) + N(x, y)\, y'\,, \tag{3.16}$$

the functional

$$J[y] \;=\; \int_{a}^{b} f(x, y, y') \; dx \tag{3.17}$$

may be rewritten

$$J[y] \;=\; \int_{a}^{b} M(x, y) \; dx \;+\; N(x, y) \; dy\,. \tag{3.18}$$

Now, if

$$M(x,y) \;=\; \frac{\partial S}{\partial x}\,, \quad N(x,y) \;=\; \frac{\partial S}{\partial y} \tag{3.19}$$

for some function $S(x,y)$, then

$$J[y] \;=\; \int_a^b dS \;=\; S(b,y_b) - S(a,y_a) \tag{3.20}$$

and we have path independence.

**Example 3.3.**

Consider

$$J[y] \;=\; \int_0^1 (y + xy')\; dx \tag{3.21}$$

subject to the boundary conditions

$$y(0) \;=\; 0\,, \quad y(1) \;=\; 1\,. \tag{3.22}$$

The Euler–Lagrange equation reduces to the identity

$$1 \;=\; 1 \tag{3.23}$$

and the integral is independent of the path of integration. Indeed,

$$\int_0^1 (y + xy')\; dx \;=\; \int_{(0,0)}^{(1,1)} y\; dx \;+\; x\; dy \tag{3.24}$$

so that

$$\int_0^1 (y + xy')\; dx \;=\; \int_{(0,0)}^{(1,1)} d(xy) \;=\; xy|_{(0,0)}^{(1,1)} \;=\; 1\,. \tag{3.25}$$

We will generally stay away from degenerate functionals. Fortunately, there are more interesting cases ahead.

**3.1.2. No explicit $y$ dependence.** Let us now consider integrands that do not depend on $y$,

$$J[y] \;=\; \int_a^b f(x, y')\; dx\,. \tag{3.26}$$

In this case, the Euler–Lagrange equation reduces to

$$\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) \;=\; 0 \tag{3.27}$$

or

$$\frac{\partial f}{\partial y'} \;=\; c\,. \tag{3.28}$$

In mechanics, $y$ is then referred to as a *cyclic* or *ignorable coordinate* and the first integral corresponds to conservation of momentum.

**3.1.3. No explicit $x$ dependence.** Finally, let us consider integrands that do not depend on $x$,

$$J[y] \;=\; \int_a^b f(y, y')\; dx\,. \tag{3.29}$$

In this case, the Euler–Lagrange equation reduces to

$$f_y - \frac{d}{dx} f_{y'} \;=\; f_y - f_{y'y}y' - f_{y'y'}y'' \;=\; 0\,. \tag{3.30}$$

Now, observe that

$$\begin{aligned} \frac{d}{dx}(f - y' f_{y'}) \;&=\; f_y y' + f_{y'}y'' - y' f_{y'y}y' - y' f_{y'y'}y'' - y'' f_{y'} \\ &=\; f_y y' - y' f_{yy'}y' - y' f_{y'y'}y'' \\ &=\; (f_y - f_{y'y}y' - f_{y'y'}y'')\, y'\,. \end{aligned} \tag{3.31}$$

Thus

$$\frac{d}{dx}(f - y' f_{y'}) \;=\; 0 \tag{3.32}$$

is equivalent to the Euler–Lagrange equation, *as long as* $y' \neq 0$. This, in turn, implies that the Euler–Lagrange equation has the first integral

$$f - y' f_{y'} \;=\; c\,. \tag{3.33}$$

In mechanics, this leads, in the simplest cases, to conservation of energy. More generally, we will get conservation of the Hamiltonian. More on this later.

Many well-known problems in the calculus of variations have integrands that do not depend explicitly on the independent variable $x$. Let us consider two examples.

## 3.2. Case study: Minimal surface of revolution

Let us return to the problem of minimizing the area

$$J[y] \;=\; 2\pi \int_a^b y(x)\,\sqrt{1+y'^2}\;dx \tag{3.34}$$

of a surface of revolution subject to the boundary conditions

$$y(a) \;=\; y_a\,, \quad y(b) \;=\; y_b \tag{3.35}$$

in the plane.

Since $x$ is not explicitly present in the integrand, we expect the first integral

$$f - y' f_{y'} \;=\; \alpha\,. \tag{3.36}$$

For this problem, this first integral gives

$$y\sqrt{1+y'^2} - y'\frac{yy'}{\sqrt{1+y'^2}} \;=\; \alpha\,, \tag{3.37}$$

where I have taken the liberty of absorbing the $2\pi$ into the $\alpha$. This last equation simplifies to

$$\frac{y}{\sqrt{1+y'^2}} \;=\; \alpha\,. \tag{3.38}$$

For $\alpha = 0$, $y = 0$, but $y = 0$ does not satisfy $y' \neq 0$. Indeed, $y = 0$ is a spurious solution: you can easily show that $y = 0$ does not satisfy the fully expanded Euler–Lagrange equation. For $y(x) > 0$ and $\alpha > 0$,

$$y \;=\; \alpha\sqrt{1+y'^2} \tag{3.39}$$

and there are now several ways to proceed.

The most direct approach is to square both sides of this equation and to rearrange terms,

$$y'^2 = \frac{1}{\alpha^2}(y^2 - \alpha^2)\,, \tag{3.40}$$

so that

$$y' = \pm\frac{1}{\alpha}\sqrt{y^2 - \alpha^2}\,. \tag{3.41}$$

This last equation is separable. There are two roots on the right-hand side, but they both, ultimately, lead to the same solution. Let us take the positive root on the right-hand side. After separating variables, we now have

$$\int \frac{1}{\sqrt{y^2 - \alpha^2}}\, dy = \frac{1}{\alpha}\int dx\,. \tag{3.42}$$

If we let

$$y = \alpha \cosh u \tag{3.43}$$

and use the facts that

$$\cosh^2 u - 1 = \sinh^2 u \tag{3.44}$$

and that

$$\frac{d}{du}\cosh u = \sinh u\,, \tag{3.45}$$

we obtain the simple integrals

$$\int du = \frac{1}{\alpha}\int dx \tag{3.46}$$

and the solution

$$u = \frac{1}{\alpha}(x - \beta)\,. \tag{3.47}$$

Our solution now reduces to the *catenary*

$$y(x) = \alpha \cosh\left(\frac{x - \beta}{\alpha}\right)\,. \tag{3.48}$$

Since we are now attuned to the usefulness of hyperbolic functions, let me point out that we could also have made the substitution

$$y' = \sinh u \tag{3.49}$$

directly into equation (3.39). This then implies that

$$y = \alpha \cosh u\,. \tag{3.50}$$

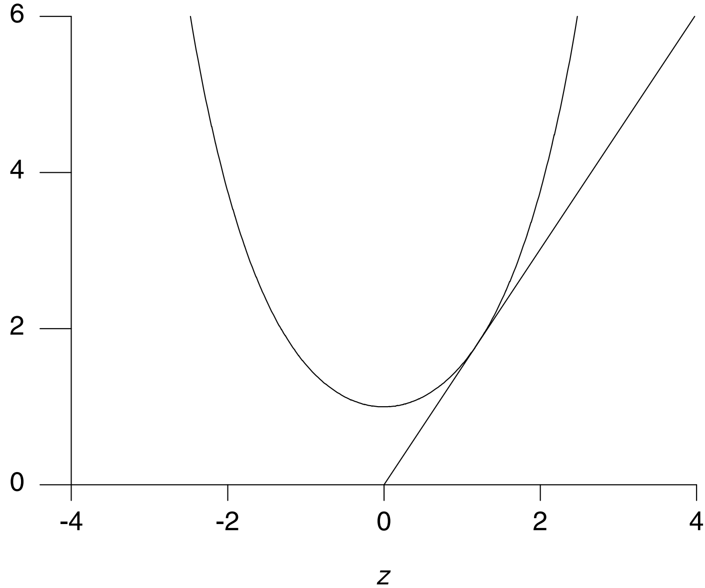



**Figure 3.1.** Solving a transcendental equation

In addition,

$$dx = \frac{1}{y'}\,dy = \frac{\alpha \sinh u\,du}{\sinh u} = \alpha\,du \tag{3.51}$$

so that

$$x = \alpha\,u + \beta\,. \tag{3.52}$$

Eliminating $u$ from parametric equations (3.50) and (3.52) once again yields equation (3.48) for our catenary.

The constants of integration $\alpha$ and $\beta$ are determined by the boundary conditions, but it is not immediately clear how many solutions, if any, arise for a given set of boundary conditions.

To give an idea of what may happen, in the simplest possible context, let us consider the symmetric boundary conditions

$$y(-h) = k\,, \quad y(h) = k\,. \tag{3.53}$$

This forces

$$k \;=\; \alpha \cosh\left(\frac{-h-\beta}{\alpha}\right) \;=\; \alpha \cosh\left(\frac{h-\beta}{\alpha}\right) \qquad (3.54)$$

and we may quickly conclude that $\beta = 0$ and that

$$\alpha \cosh\frac{h}{\alpha} \;=\; k\,. \qquad (3.55)$$

If we let

$$z \;\equiv\; \frac{h}{\alpha}\,, \qquad (3.56)$$

we may now write

$$\cosh z \;=\; mz\,, \qquad (3.57)$$

where

$$m \;\equiv\; \frac{k}{h}\,. \qquad (3.58)$$

We can draw a catenary between our symmetric boundary conditions if we can find real roots of equation (3.57). This equation is transcendental, but we can solve it graphically (see Figure 3.1). All we need to do is to plot the left-hand side and the right-hand side of equation (3.57) as functions of $z$ and to look for intersections.

For $m = k/h$ sufficiently large, we have two roots and *two* catenaries that satisfy our boundary conditions. If our boundaries are far enough apart, there are no catenary solutions. These two extremes are separated by a critical case, $m = m_c$, $z = z_c$, that corresponds to the double root

$$\cosh z_c \;=\; m_c\, z_c\,, \quad \sinh z_c \;=\; m_c\,. \qquad (3.59)$$

We can eliminate $m_c$ between these two equations to get

$$\tanh z_c \;=\; \frac{1}{z_c}\,, \qquad (3.60)$$

which has the solution

$$z_c \;\approx\; 1.199679\,. \qquad (3.61)$$

It now follows that

$$m_c \;\approx\; \sinh(1.199679) \;\approx\; 1.508880\,. \qquad (3.62)$$

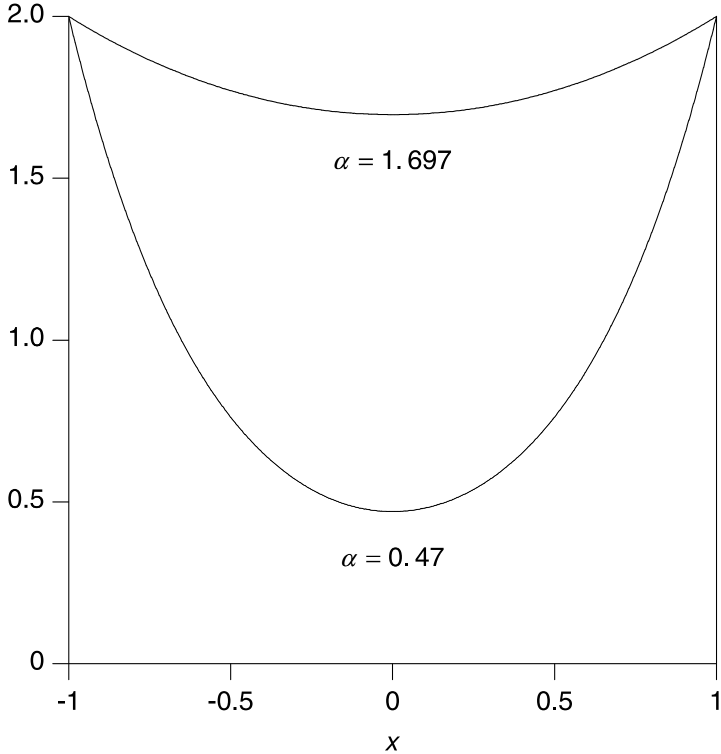



**Figure 3.2.** Two catenaries

For $m > m_c$ two catenaries solve the boundary conditions. One of these will be shallow and one of these will be deep. For example, for

$$y(-1) = 2\,, \quad y(1) = 2\,, \tag{3.63}$$

we have $m = 2$. The transcendental equation

$$\cosh z = mz \tag{3.64}$$

now has two roots,

$$z_1 \approx 0.58939\,, \quad z_2 \approx 2.1268 \tag{3.65}$$

with

$$\alpha_1 \approx 1.697\,, \quad \alpha_2 \approx 0.47\,. \tag{3.66}$$

Figure 3.2 shows the two catenaries.

We will see later that the deeper of the two catenaries is a spurious stationary point that is neither a maximum nor a minimum. We will also see that the shallower of the two catenaries is a relative minimum.

There is another way to view matters (see Figure 3.3). Let us take our first boundary condition to be

$$y(0) \;=\; 1\,. \tag{3.67}$$

The catenary

$$y(x) \;=\; \alpha \cosh\left(\frac{x-\beta}{\alpha}\right) \tag{3.68}$$

must now satisfy

$$1 \;=\; \alpha \cosh\frac{\beta}{\alpha}\,. \tag{3.69}$$

Let us denote

$$\lambda \;=\; \frac{\beta}{\alpha} \tag{3.70}$$

so that

$$\alpha \;=\; \frac{1}{\cosh\lambda}\,. \tag{3.71}$$

Our two-parameter family of catenaries now reduces to the one-parameter family

$$y(x,\lambda) \;=\; \frac{\cosh\left(x\cosh\lambda-\lambda\right)}{\cosh\lambda}\,. \tag{3.72}$$

Since

$$y'(x,\lambda) \;=\; \sinh\left(x\cosh\lambda-\lambda\right) \tag{3.73}$$

and

$$y'(0,\lambda) \;=\; -\sinh\lambda\,, \tag{3.74}$$

we can clearly obtain all possible initial slopes if we let $\lambda$ run through all real numbers.

As we vary $\lambda$, we obtain a one-parameter family or “pencil” of catenaries emanating from $(1,0)$. The members of this family form an envelope that passes through the origin. For some parametric representations of this envelope, see Kimball (1952). Every point $P$ above this envelope is visited by two different catenaries. One of the two catenaries touches the envelope before passing on to $P$. We will

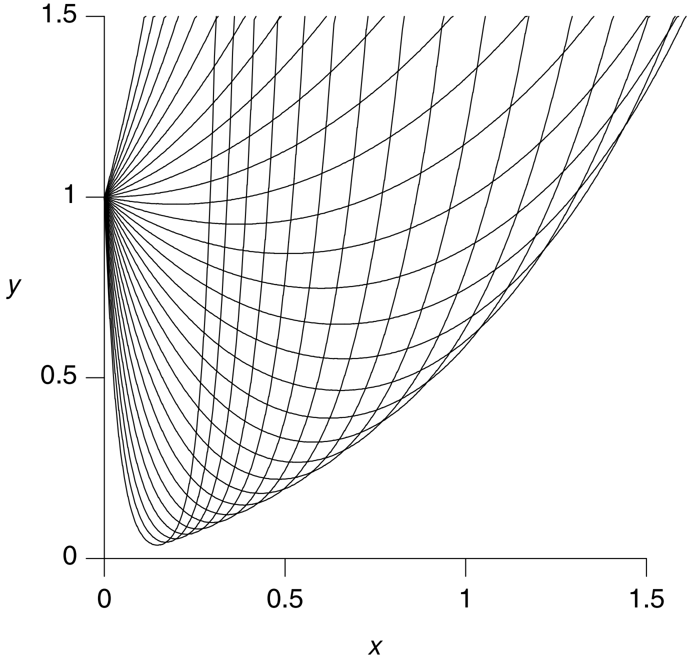



**Figure 3.3.** One-parameter family of catenaries

see later that this catenary is not a minimizing curve. The other catenary, which does not touch the envelope, is a relative minimum.

Points below the envelope cannot be reached by a catenary. When the point $P$ is on or below the envelope, there is no continuously differentiable curve of the form $y = y(x)$ that generates a relative minimum surface of revolution. The minimal surface of revolution consists, instead, of two boundary disks (corresponding to the boundary conditions) and the segment of the $x$-axis between them. This solution is called the *Goldschmidt solution* in honor of (Charles Wolfgang) Benjamin Goldschmidt (1831) and his discovery of this solution. This Goldschmidt solution coexists with our two catenaries when point $P$ is above the envelope. The Goldschmidt solution may or may not be the global minimum in these cases.

## 3.3. Case study: The brachistochrone

Let us reexamine the brachistochrone problem. The problem was to minimize the travel time of a mass moving under its own weight between two points. That is, we want to minimize the integral

$$T = \int_a^b \frac{1}{v}\, ds = \int_a^b \frac{1}{v} \sqrt{1 + y'^2}\, dx \tag{3.75}$$

subject to the boundary conditions

$$y(a) = y_a\,, \quad y(b) = y_b\,. \tag{3.76}$$

Remember, this is a conservative system. For a particle starting from rest,

$$\frac{1}{2}\, mv^2 + mgy = mgy_a \tag{3.77}$$

and

$$v = \sqrt{2g(y_a - y)}\,. \tag{3.78}$$

We are thus left with the problem of minimizing

$$J[y] = \frac{1}{\sqrt{2g}} \int_a^b \sqrt{\frac{1 + y'^2}{y_a - y}}\, dx \tag{3.79}$$

subject to the boundary conditions.

Let's make our lives a little easier by letting

$$z \equiv y_a - y \tag{3.80}$$

so that

$$J[z] = \frac{1}{\sqrt{2g}} \int_a^b \sqrt{\frac{1 + z'^2}{z}}\, dx\,. \tag{3.81}$$

Since there is no explicit $x$ dependence in this integral, the Euler–Lagrange equation has the first integral

$$f - z' \frac{\partial f}{\partial z'} = \alpha \tag{3.82}$$

or

$$\frac{\sqrt{1 + z'^2}}{\sqrt{z}} - \frac{1}{\sqrt{z}} \frac{z'^2}{\sqrt{1 + z'^2}} = \alpha\,. \tag{3.83}$$

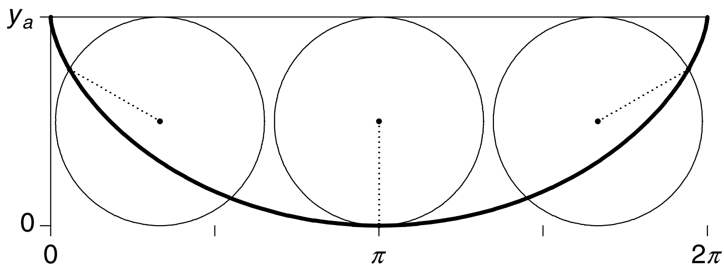



**Figure 3.4.** Cycloid for $R = \frac{1}{2}y_a$ and $a = 0$

This last equation simplifies to

$$\frac{1}{\sqrt{z[1+(z')^2]}} = \alpha\,. \tag{3.84}$$

Solving for $z'$, we find that

$$\frac{dz}{dx} = \sqrt{\frac{1-\alpha^2 z}{\alpha^2 z}}\,. \tag{3.85}$$

This last differential equation is separable,

$$dx = \sqrt{\frac{\alpha^2 z}{1-\alpha^2 z}}\,dz\,. \tag{3.86}$$

Let

$$z = \frac{1}{\alpha^2}\sin^2\theta \tag{3.87}$$

so that

$$dz = \frac{2}{\alpha^2}\sin\theta\,\cos\theta\,d\theta\,. \tag{3.88}$$

Our separated differential equation now reduces to

$$dx = \frac{2}{\alpha^2}\sqrt{\frac{\sin^2\theta}{\cos^2\theta}}\,\sin\theta\,\cos\theta\,d\theta \tag{3.89}$$

or

$$\alpha^2\,dx = 2\sin^2\theta\,d\theta = (1-\cos 2\theta)\,d\theta\,. \tag{3.90}$$

It follows that

$$\alpha^2\,x = \theta - \frac{1}{2}\sin 2\theta + \beta\,. \tag{3.91}$$

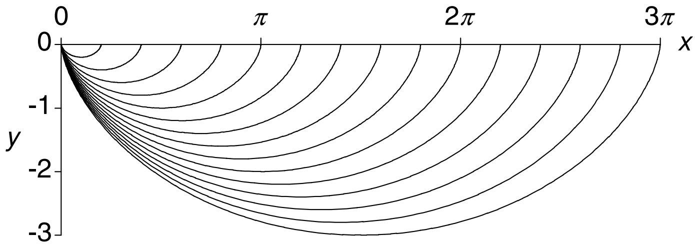



**Figure 3.5.** One-parameter family of cycloids

Now, remember that

$$y \;=\; y_a - z \tag{3.92}$$

so that

$$y \;=\; y_a - \frac{1}{\alpha^2}\sin^2\theta \;=\; y_a - \frac{1}{2\alpha^2}(1-\cos 2\theta)\,. \tag{3.93}$$

If we let

$$R \equiv \frac{1}{2\alpha^2}\,, \quad \phi \equiv 2\theta\,, \quad a = \frac{\beta}{\alpha^2}\,, \tag{3.94}$$

we may write our solution in the parametric form

$$x(\phi) \;=\; a + R\,(\phi - \sin\phi)\,, \tag{3.95}$$

$$y(\phi) \;=\; y_a - R\,(1 - \cos\phi)\,. \tag{3.96}$$

This is the trace of a circle of radius $R$ rolling on the underside of the line $y = y_a$ (see Figure 3.4); the equation is that of a *cycloid.*

If we take

$$x(0) \;=\; 0, \quad y(0) \;=\; 0 \tag{3.97}$$

as our left boundary condition, we are left with a one-parameter family of cycloids,

$$x \;=\; R\,(\phi - \sin\phi)\,, \quad y \;=\; -R\,(1-\cos\phi)\,, \tag{3.98}$$

that form a field of extremals (see Figure 3.5). Each extremal extends along the $x$-axis $\pi$ times its maximum depth. There is one (and only one) cycloid that passes through the boundary points $A = (0,\,0)$ and $B$ with no cusp between $A$ and $B$.

## 3.4. Geodesics

In Chapter 1, we saw that we can represent a surface as a vector,

$$\mathbf{r}(u,v) \;=\; x(u,v)\;\mathbf{i} + y(u,v)\;\mathbf{j} + z(u,v)\;\mathbf{k}\,. \tag{3.99}$$

Finding the shortest path between two points, $A = (u_a, v_a)$ and $B = (u_b, v_b)$, on this surface was then achieved by minimizing the integral

$$s \;=\; \int_{u_a}^{u_b} \sqrt{E + 2\,F\left(\frac{dv}{du}\right) + G\left(\frac{dv}{du}\right)^2}\;du \tag{3.100}$$

subject to the boundary conditions

$$v(u_a) \;=\; v_a\,, \quad v(u_b) \;=\; v_b\,. \tag{3.101}$$

Here,

$$E(u,v) \;=\; \frac{\partial \mathbf{r}}{\partial u}\cdot\frac{\partial \mathbf{r}}{\partial u}\,, \quad F(u,v) \;=\; \frac{\partial \mathbf{r}}{\partial u}\cdot\frac{\partial \mathbf{r}}{\partial v}\,, \tag{3.102}$$
$$G(u,v) \;=\; \frac{\partial \mathbf{r}}{\partial v}\cdot\frac{\partial \mathbf{r}}{\partial v}$$

are the first-order fundamental quantities.

For general surfaces, the resulting Euler–Lagrange equation,

$$\frac{E_v + 2F_v v' + G_v (v')^2}{2\sqrt{E + 2Fv' + G(v')^2}} - \frac{d}{du}\left[\frac{F + Gv'}{\sqrt{E + 2Fv' + G(v')^2}}\right] \;=\; 0\,, \tag{3.103}$$

is quite awful. Fortunately, this Euler–Lagrange equation does simplify for some surfaces.

The most important special case is that of a *surface of revolution,* which is obtained by rotating a planar curve about an axis. If we choose the $z$-axis as the axis of revolution, we may represent the surface of revolution by the vector

$$\mathbf{r}(u,v) \;=\; f(u)\,\cos v\;\mathbf{i} + f(u)\,\sin v\;\mathbf{j} + h(u)\;\mathbf{k}\,. \tag{3.104}$$

Varying $u$ while holding $v$ constant yields a *meridian* of the surface. Varying $v$ while holding $u$ constant yields a *circle of latitude* for the surface.

For this surface of revolution, the first-order fundamental quantities are

$$E \;=\; [f'(u)]^2 + [h'(u)]^2\,, \quad F \;=\; 0\,, \quad G \;=\; [f(u)]^2\,. \tag{3.105}$$

These quantities do not depend on the dependent variable $v$. In addition, $F = 0$, which implies that the meridians and circles of latitude intersect orthogonally. As a result, our Euler–Lagrange equation simplifies to

$$\frac{d}{du}\left[\frac{G\,v'}{\sqrt{E+G\,(v')^2}}\right] \;=\; 0\,. \tag{3.106}$$

We thus gain the first integral

$$\frac{G\,v'}{\sqrt{E+G\,(v')^2}} \;=\; c_1\,. \tag{3.107}$$

If we solve for $v'(u)$ and integrate, it now follows that

$$v(u) \;=\; c_1 \int \frac{\sqrt{E}}{\sqrt{G^2 - c_1^2 G}}\; du \,+\, c_2\,. \tag{3.108}$$

On occasion, you may encounter problems in which $F = 0$ and the *independent* variable $u$ is missing from the first-order fundamental quantities. In that case, the Euler–Lagrange equation reduces to

$$\sqrt{E+G(v')^2} - \frac{G(v')^2}{\sqrt{E+G(v')^2}} \;=\; c_1\,. \tag{3.109}$$

After some simplification, it now follows that

$$u(v) \;=\; c_1 \int \frac{\sqrt{G}}{\sqrt{E^2 - c_1^2 E}}\; dv \,+\, c_2\,. \tag{3.110}$$

You will find several problems involving geodesics at the end of this chapter.

## 3.5. Recommended reading

We will consider the catenoid in greater detail in Chapter 6. In the meantime, see Durand (1981) and Ben Amar et al. (1998) for information about the stability and oscillation of soap films.

A number of scientists have investigated the collapse of catenoidal soap-film bridges experimentally. See, in particular, the work of Cryer and Steen (1992), Chen and Steen (1997), Robinson and Steen (2001), and Müller and Stannarius (2006).

Although we typically associate soap films with minimal surfaces such as the catenoid, Criado and Alamo (2010) have shown that other variational problems, including the brachistochrone, can also be solved using soap films.

The classical brachistochrone has a point mass sliding along a curve. One can also consider brachistochrones for rigid bodies, such as disks, cylinders, and spheres, rolling down a curve. See Rodgers (1946), Yang et al. (1987), Akulenko (2009), and Legeza (2010) for details.

Since the Euler–Lagrange equation for a geodesic on a general surface is often analytically intractable, scientists commonly use numerical methods to find geodesics. Patrikalakis and Maekawa (2002) provide a nice overview of numerical techniques for finding geodesics.

## 3.6. Exercises

**3.6.1. Finding extremals.** Find extremals for the following functionals (Elsgolc, 1961). All are well-known geometric curve. For each extremal, name, draw, or describe the curve.

(a)

$$F[y(x)] = \int_a^b \frac{\sqrt{1+y'^2}}{y}\, dx. \tag{3.111}$$

(b)

$$F[y(x)] = \int_a^b (y^2 + 2xyy')\, dx, \tag{3.112}$$

$$y(a) = y_a \quad \text{and} \quad y(b) = y_b. \tag{3.113}$$

(c)

$$F[y(x)] = \int_0^1 (xy + y^2 - 2y^2y')\, dx, \tag{3.114}$$

$$y(0) = 1 \quad \text{and} \quad y(1) = 2. \tag{3.115}$$

(d)

$$F[y(x)] = \int_a^b \sqrt{y(1+y'^2)}\, dx. \tag{3.116}$$

(e)

$$F[y(x)] = \int_a^b y'(1 + x^2 y')\, dx. \tag{3.117}$$

(f)

$$F[y(x)] = \int_a^b (y'^2 + 2yy' - 16y^2)\, dx. \tag{3.118}$$

**3.6.2. An autonomous equation.** A differential equation is *autonomous* if it does not explicitly depend on or contain the independent variables.

(a) Show that the Euler–Lagrange equation for the minimal surface of revolution problem reduces to the autonomous, nonlinear differential equation

$$1 + \left(\frac{dy}{dx}\right)^2 - y\frac{d^2y}{dx^2} = 0\,. \tag{3.119}$$

(b) Write $y'(x)$ as a function of $y$,

$$y' = p(y)\,, \tag{3.120}$$

so that

$$y'' = \frac{dp}{dy}\frac{dy}{dx} = p\,\frac{dp}{dy}\,. \tag{3.121}$$

Use this substitution to rederive the first-order equation

$$y'^2 = \frac{1}{\alpha^2}(y^2 - \alpha^2)\,. \tag{3.122}$$

**3.6.3. A linear equation.** Differentiate the first-order equation

$$y'^2 = \frac{1}{\alpha^2}(y^2 - \alpha^2) \tag{3.123}$$

(from the previous problem) with respect to $x$ to derive a linear, second-order differential equation. Solve for the general solution to this ODE and show that it contains three arbitrary constants. Use equation (3.123) to eliminate one constant and rederive the catenary of equation (3.48).

**3.6.4. Brachistochrones on a cylinder.** Determine the extremals for the brachistochrone on the cylinder (Vujanovic and Jones, 1989). Find the minimum-time curve of descent of a particle, under the influence of gravity, on a vertical circular cylinder of fixed radius $r$. Assume that the particle starts from rest, that the force of friction between the particle and the cylinder is negligible, and that we may describe the location of the particle in terms of the cylindrical coordinates $\theta$ and $z$.

(a) Write out the functional for this problem with $z(\theta)$ as the dependent variable and $\theta$ as the independent variable.

(b) Determine the extremals for this problem by solving the Euler-Lagrange equation for this problem.

(c) What happens to your solution as you "unroll" your cylinder?

**3.6.5. Terrestrial brachistochrones.** Consider the problem of minimizing the travel time through the earth,

$$T = \sqrt{\frac{R}{g}} \int_{\theta_a}^{\theta_b} \sqrt{\frac{\left(\frac{dr}{d\theta}\right)^2 + r^2}{R^2 - r^2}}\, d\theta\,. \tag{3.124}$$

(See Figure 1.3.) Determine the extremals for this problem, $\theta(r)$, by solving the Euler–Lagrange equation for this problem.

**3.6.6. Geodesics on a right circular cylinder.** Let $S$ be the right circular cylinder

$$\mathbf{r}(\theta, z) = a \cos\theta\, \mathbf{i} + a \sin\theta\, \mathbf{j} + z\, \mathbf{k} \tag{3.125}$$

of radius $a$, with $0 \le \theta \le 2\pi$ and $-\infty < z < \infty$. A curve $\gamma$, on the surface of the cylinder, is given by a function, $z = z(\theta)$, that relates $z$ and $\theta$ along $\gamma$. We wish to find the geodesic that minimizes the arc-length integral between two points on the surface of the cylinder.

(a) Write down the arc-length integral for the curve $\gamma$.

(b) What is the Euler–Lagrange equation for this integral?

(c) Solve this differential equation and show that your geodesic curves are arcs of circular helices.

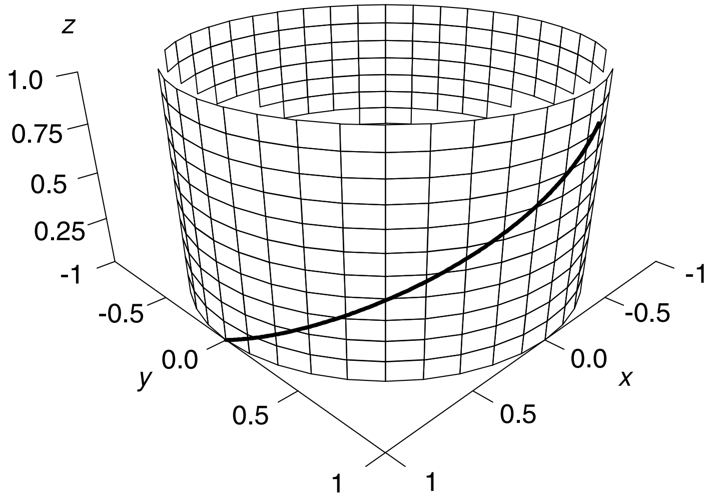



**Figure 3.6.** Geodesic on a cylinder

**3.6.7. Geodesics on a right circular cone.** Let $S$ be the right circular cone

$$\mathbf{r}(r,\theta) \;=\; r\cos\theta\,\mathbf{i} + r\sin\theta\,\mathbf{j} + mr\,\mathbf{k}\,, \tag{3.126}$$

with $0 \le \theta \le 2\pi$, $r \ge 0$, and $m$ a constant (the slope of the cone). A curve $\gamma$, on the surface of the cone, is given by a function, $r = r(\theta)$, that relates $r$ and $\theta$ along $\gamma$. We wish to find the geodesic that minimizes the arc-length integral between two points on the surface of the cone.

(a) Write down the arc-length integral for the curve $\gamma$.

(b) Show that if $\gamma$ is a geodesic, $r = r(\theta)$ must satisfy the differential equation

$$(1+m^2)\,\alpha^2\left(\frac{dr}{d\theta}\right)^2 \;=\; r^2\,(r^2-\alpha^2) \tag{3.127}$$

for some suitable constant $\alpha$ $(0 \le \alpha \le r)$.

(c) Integrate this differential equation to get the equation

$$r(\theta) = \frac{\alpha}{\cos\left[\frac{(\theta+\beta)}{\sqrt{1+m^2}}\right]} \tag{3.128}$$

for a geodesic curve that winds around and up the cone (with nonconstant radius). Here $\alpha$ and $\beta$ are appropriate constants of integration.

**3.6.8. Geodesics on a sphere.** Let $S$ be a sphere,

$$\mathbf{r}(\theta, \phi) = R \sin\theta \cos\phi \, \mathbf{i} + R \sin\theta \sin\phi \, \mathbf{j} + R \cos\theta \, \mathbf{k}, \tag{3.129}$$

of radius $R$, with $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$. A curve $\gamma$, on the surface of the sphere, is given by a function, $\phi = \phi(\theta)$, that relates $\phi$ and $\theta$ along $\gamma$. We wish to find the geodesic that minimizes the arc-length integral between two points on the surface of the sphere.

(a) What is the Euler–Lagrange equation for this problem?

(b) Solve your Euler–Lagrange equation for this problem and show that the extremals are arcs of great circles.

Chapter 4

# Basic Generalizations

## 4.1. Higher-order derivatives

We will now consider some simple generalizations of our basic theory. We will start with functionals that contain a second derivative.

Consider the problem of maximizing or minimizing the integral

$$J[y] \;=\; \int_a^b f(x, y(x), y'(x), y''(x))\; dx \tag{4.1}$$

subject to the boundary conditions

$$y(a) \;=\; y_a\,, \quad y(b) \;=\; y_b\,, \quad y'(a) \;=\; y'_a\,, \quad y'(b) \;=\; y'_b\,. \tag{4.2}$$

This is similar to our standard problem in Chapter 2, except that the functional now contains a second derivative. The boundary conditions, in turn, prescribe both $y(x)$ and $y'(x)$ at the endpoints.

We will proceed in the usual manner. We will assume that we have a solution, $y = \hat{y}(x)$, and add a small weak variation, $h(x) = \epsilon\eta(x)$, to this solution,

$$y(x) \;=\; \hat{y}(x) + \epsilon\,\eta(x)\,. \tag{4.3}$$

If this new function is to satisfy our boundary conditions, $\eta(x)$ and its derivative must both vanish at the endpoints,

$$\eta(a) \;=\; 0\,, \quad \eta(b) \;=\; 0\,, \quad \eta'(a) \;=\; 0\,, \quad \eta'(b) \;=\; 0\,. \tag{4.4}$$

The total variation is now

$$\Delta J \;=\; \int_a^b f(x, \hat{y} + \epsilon\eta, \hat{y}' + \epsilon\eta', \hat{y}'' + \epsilon\eta'') - f(x, \hat{y}, \hat{y}', \hat{y}'')\; dx \qquad (4.5)$$

and, if we expand the integrand in a Taylor series in $\epsilon$ and keep terms of order $\epsilon$, the first variation is just

$$\delta J \;=\; \epsilon \int_a^b \left( \frac{\partial f}{\partial y}\eta + \frac{\partial f}{\partial y'}\eta' + \frac{\partial f}{\partial y''}\eta'' \right) dx\,. \qquad (4.6)$$

We may now simplify the first variation by integrating by parts. Since

$$\int_a^b \eta' \frac{\partial f}{\partial y'}\; dx \;=\; \eta \frac{\partial f}{\partial y'}\Bigg|_{x=a}^{x=b} - \int_a^b \eta \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) dx \qquad (4.7)$$

and

$$\int_a^b \eta'' \frac{\partial f}{\partial y''}\; dx \;=\; \eta' \frac{\partial f}{\partial y''}\Bigg|_{x=a}^{x=b} - \int_a^b \eta' \frac{d}{dx}\left(\frac{\partial f}{\partial y''}\right) dx \qquad (4.8)$$

$$= \left[\eta' \frac{\partial f}{\partial y''} - \eta \frac{d}{dx}\left(\frac{\partial f}{\partial y''}\right)\right]_{x=a}^{x=b} + \int_a^b \eta \frac{d^2}{dx^2}\left(\frac{\partial f}{\partial y''}\right) dx\,,$$

the first variation reduces to

$$\delta J \;=\; \epsilon \int_a^b \eta(x) \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial f}{\partial y''}\right)\right] dx \qquad (4.9)$$

$$+\, \epsilon \left[\left(\frac{\partial f}{\partial y'} - \frac{d}{dx}\frac{\partial f}{\partial y''}\right)\eta\right]_{x=a}^{x=b} + \epsilon\, \frac{\partial f}{\partial y''}\eta'\Bigg|_{x=a}^{x=b}.$$

We have seen that we must have $\delta J = 0$ for a relative maximum or minimum. If $y(x)$ and $y'(x)$ are prescribed at $x = a$ and $x = b$, then $\eta(x)$ and its derivative vanish at those endpoints, and the last two terms in equation (4.9) vanish. It now follows that $\hat{y}(x)$ must satisfy the fourth-order equation

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial f}{\partial y''}\right) \;=\; 0\,. \qquad (4.10)$$

Equation (4.10) simplifies for a number of special cases. These cases include degenerate functionals and first integrals. See Exercises 4.10.1, 4.10.2, and 4.10.3 for details.

For many applications (see, for example, the next section), $y(x)$ and $y'(x)$ are left unprescribed at one (or both) endpoints. The functions $\eta(x)$ and $\eta'(x)$ need not vanish at those endpoints. To force our first variation to vanish, we must then augment our fourth-order Euler–Lagrange equation with "natural" boundary conditions. If $y$ is unprescribed at an endpoint, we require that

$$\frac{\partial f}{\partial y'} - \frac{d}{dx}\left(\frac{\partial f}{\partial y''}\right) = 0 \tag{4.11}$$

at that endpoint. If $y'$ is unprescribed at an endpoint, we require that

$$\frac{\partial f}{\partial y''} = 0 \tag{4.12}$$

at that endpoint. We will discuss natural boundary conditions and other endpoint conditions more fully in Chapter 9.

The arguments that led to equation (4.10) can also be extended to integrals,

$$J = \int_a^b f(x, y(x), y'(x), \ldots, y^{(n)}(x))\, dx\,, \tag{4.13}$$

that contain derivatives of order $n$. I leave it to you (see Exercise 4.10.4) to show that $\hat{y}(x)$ must now satisfy the *Euler–Poisson equation*

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) + \cdots + (-1)^n \frac{d^n}{dx^n}\left[\frac{\partial f}{\partial y^{(n)}}\right] = 0\,, \tag{4.14}$$

a differential equation of order $2n$.

## 4.2. Case study: The cantilever beam

The *principle of minimum (total) potential energy* is an important tool in solid mechanics. This principle is used for elastic systems that experience conservative internal and external forces and that lie, at rest, in stable equilibrium. The principle states that the total potential energy for such a system is a minimum with respect to all small displacements that satisfy the given boundary conditions. This principle is used to derive the shape of an elastic body.

To be sure that we are speaking the same language, remember that an elastic body is a deformable solid body whose deformations disappear after the forces that caused the deformations have been removed. Elastic bodies return to their original shape. In linear theories of elasticity, we assume that deformations and strains are small and that stress and strain are linearly related. A conservative force, in turn, is one that always does the same work moving a particle between two given points, irrespective of the path. Conservative forces conserve mechanical energy and can be derived from scalar potentials.

In a moment, we will derive the equation for the shape of a cantilever beam (fixed at one end and free at the other) under a uniform load. In this example, the external force is a familiar one: gravity. The internal forces arise from strains within the beam. The total potential energy has two components: elastic strain energy and gravitational potential energy. We will start with the strains and the strain energy within a bent beam.

Consider a straight beam of uniform cross section (see Figure 4.1). We will subject this beam to pure bending. We will assume that this bending occurs in the $(x, y)$ plane, that there is no lateral shearing stress, and that each transverse section of the bar, initially plane, remains plane and perpendicular to the longitudinal fibers of the bar. These are standard assumptions for Bernoulli–Euler beam theory.

As the beam bends (see Figure 4.2), longitudinal fibers above the original $x$-axis shorten, while longitudinal fibers below the $x$-axis lengthen. We will assume that the center fiber, the neutral axis, remains unstrained. For convenience, let $R$ be the radius of curvature of this neutral axis.

What is the strain, after bending, of a typical fiber? Consider fiber $CG$, located a distance $u$ above the neutral axis. Before bending, this fiber is the same length as $BF$,

$$CG \; = \; BF \; = \; R\,\Delta\theta\,. \tag{4.15}$$

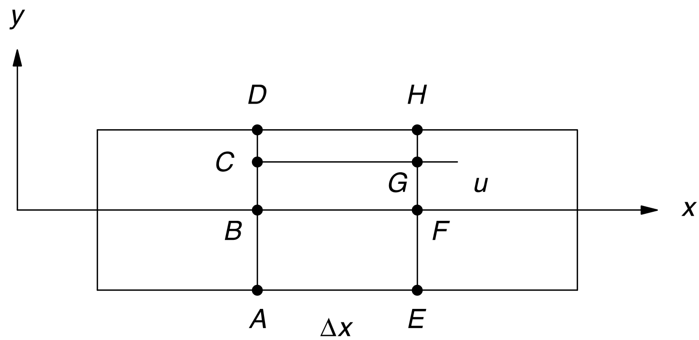



**Figure 4.1.** Straight beam

After bending, this fiber is shorter,

$$CG \;=\; (R - u)\,\Delta\theta\,. \tag{4.16}$$

The strain $\epsilon$ is simply the proportional change in length,

$$\epsilon \;=\; -\frac{u\,\Delta\theta}{R\,\Delta\theta} \;=\; -\frac{u}{R}\,. \tag{4.17}$$

The minus sign indicates that fibers above the neutral axis shorten and that fibers below the neutral axis, where $u$ is negative, lengthen.

The above formula for the strain contains the reciprocal of the radius of curvature. This is just the curvature, or the rate of change of direction of the neutral axis with arc length. The curvature can be written

$$\kappa(x) \;=\; \frac{y''}{(1+y'^2)^{3/2}} \tag{4.18}$$

(see Exercise 4.10.5), where $y(x)$ is the vertical displacement of the neutral axis. Thus, the strain is

$$\epsilon \;=\; -\,\frac{u\,y''}{(1+y'^2)^{3/2}}\,. \tag{4.19}$$

For a linearly elastic beam, Hooke's law is valid and the stress in fiber $CG$ is

$$\sigma \;=\; E\,\epsilon \tag{4.20}$$

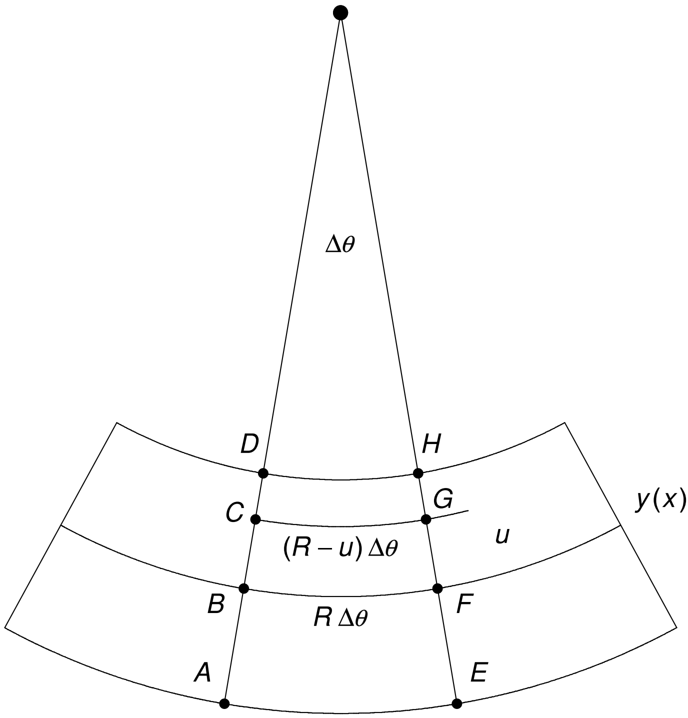



**Figure 4.2.** Bent beam

where $E$ is Young's modulus. The stress in the fibers engenders a strain energy density,

$$U_0 = \int_0^\epsilon \sigma \, d\epsilon = \int_0^\epsilon E\,\epsilon \, d\epsilon = \frac{E}{2}\,\epsilon^2 = \frac{E}{2}\,\frac{(u\,y'')^2}{(1+y'^2)^3}\,. \tag{4.21}$$

Integrating this strain energy density over the volume of the bent beam, in turn, gives us the total strain energy,

$$U = \int_V U_0 \, dV = \int_0^l \int_A U_0 \, dA\,ds\,. \tag{4.22}$$

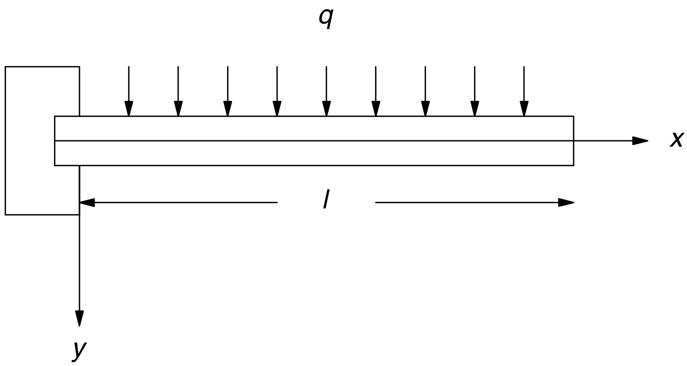



**Figure 4.3.** Cantilever beam

Substituting $U_0$ and

$$ds = \sqrt{1 + y'^2}\, dx \tag{4.23}$$

now gives us

$$U = \frac{EI}{2} \int_0^b \frac{y''^2}{(1 + y'^2)^{5/2}}\, dx\,, \tag{4.24}$$

where

$$I = \int_A u^2\, dA \tag{4.25}$$

is the moment of inertia of the cross section of the beam with respect to the centroidal $z$-axis. The limit of integration $b$ is, to a first approximation, equal to $l$.

Finally, consider a uniform elastic cantilever beam (fixed at one end and free at the other) under a uniform load (see Figure 4.3). For convenience, we will now consider the $y$-axis to be in the downward direction, so that we have positive deflections. You may think of the cantilever beam as a high-dive board with someone lying on the board.

We have seen that the bending of the beam (about the neutral line of the beam) results in the strain energy

$$U = \frac{EI}{2} \int_0^b \frac{y''^2}{(1+y'^2)^{5/2}}\, dx\,, \tag{4.26}$$

where $y(x)$ is the deflection of the beam, $E$ is Young's modulus, and $I$ is the moment of inertia about the centroidal axis. The loss in potential energy due to the external, distributed load $q$ is, in turn,

$$-\int_0^l q\, y\; ds\,. \tag{4.27}$$

The total potential energy of the deformed system can thus be written as

$$V = \int_0^b \left[ \frac{EI}{2} \frac{y''^2}{(1+y'^2)^{5/2}} - q\, y\, (1+y'^2)^{1/2} \right] dx. \tag{4.28}$$

This integral is sufficiently complicated that the resulting differential equation for the extremal is also quite complicated. However, if we assume that the beam deflection is small, we can neglect second-degree terms in $y'(x)$ and write

$$V \approx \int_0^b \left( \frac{EI}{2} y''^2 - q\, y \right) dx\,. \tag{4.29}$$

The Euler–Poisson equation for this integral reduces to

$$EI\, y'''' - q = 0\,. \tag{4.30}$$

The boundary conditions at the clamped end are

$$y(0) = 0\,, \quad y'(0) = 0\,. \tag{4.31}$$

There are no predefined boundary conditions at the unsupported end. We must instead use the natural boundary conditions. The natural boundary conditions reduce to

$$EI\, y'''(b) = 0\,, \quad EI\, y''(b) = 0\,. \tag{4.32}$$

## 4.3. Multiple unknown functions

The next generalization concerns integrals that contain multiple dependent variables. In particular, consider an integral,

$$J[y_1, \ldots, y_n] \;=\; \int_a^b f(x, y_1, \ldots, y_n, y_1', \ldots, y_n')\; dx\,, \tag{4.33}$$

that contains $n$ twice continuously differentiable functions, $y_1(x), \ldots,$ $y_n(x)$, that satisfy the boundary conditions

$$y_i(a) \;=\; y_{ia}\,, \quad y_i(b) \;=\; y_{ib} \tag{4.34}$$

for $i = 1, \ldots, n$.

If $n = 2$, we will often find it convenient, in later sections, to assign distinct names to the two dependent variables, e.g., $y(x)$ and $z(x)$. If, in contrast, $n > 2$, or if we are talking about the general case, we will use subscripts for the dependent variables, as in equation (4.33). Also, in order to save effort and ink, we will often use vector notation for the dependent variables and their derivatives,

$$\mathbf{y}(x) \;=\; [y_1(x), \ldots, y_n(x)]\,, \quad \mathbf{y}'(x) \;=\; [y_1'(x), \ldots, y_n'(x)]\,, \tag{4.35}$$

and for optimal solutions and their derivatives,

$$\hat{\mathbf{y}}(x) \;=\; [\hat{y}_1(x), \ldots, \hat{y}_n(x)]\,, \quad \hat{\mathbf{y}}'(x) \;=\; [\hat{y}_1'(x), \ldots, \hat{y}_n'(x)]\,. \tag{4.36}$$

For now, we will assume that the dependent variables are independent of one another, with no a priori constraints. We will consider constraints in Chapter 5.

We will consider weak variations,

$$h_i(x) \;=\; \epsilon\, \eta_i(x)\,, \tag{4.37}$$

that vanish at $x = a$ and $x = b$,

$$\eta_i(a) \;=\; 0\,, \quad \eta_i(b) \;=\; 0\,, \tag{4.38}$$

for $i = 1, \ldots, n$. We will often use vector notation for weak variations and their derivatives.

To find necessary conditions for the functional to have an extremum, we once again consider the total variation

$$\Delta J \;=\; J[\hat{\mathbf{y}} + \mathbf{h}] - J[\hat{\mathbf{y}}] \tag{4.39}$$

relative to an assumed solution. Thus,

$$\Delta J \;=\; \int_a^b f(x, \hat{\mathbf{y}} + \epsilon\boldsymbol{\eta}, \hat{\mathbf{y}}' + \epsilon\boldsymbol{\eta}') - f(x, \hat{\mathbf{y}}, \hat{\mathbf{y}}')\; dx \tag{4.40}$$

so that

$$\Delta J \;=\; \delta J \;+\; \frac{1}{2}\delta^2 J \;+\; O(\epsilon^3) \tag{4.41}$$

where

$$\delta J \;=\; \epsilon \int_a^b \sum_{i=1}^{n} (f_{y_i}\eta_i + f_{y_i'}\eta_i')\; dx\,. \tag{4.42}$$

Since all the increments $\eta_i(x)$ are independent, we can choose any one of these increments arbitrarily (as long as the boundary conditions are satisfied) while keeping all of the other increments equal to zero. The necessary condition $\delta J = 0$ thus yields

$$\epsilon \int_a^b (f_{y_i}\eta_i + f_{y_i'}\eta_i')\; dx \;=\; 0 \tag{4.43}$$

for $i = 1, \ldots, n$. Using the fundamental lemma of the calculus of variations or, more generally, the du Bois-Reymond lemma, we obtain a system of Euler–Lagrange equations,

$$\frac{\partial f}{\partial y_i} - \frac{d}{dx}\left(\frac{\partial f}{\partial y_i'}\right) \;=\; 0 \tag{4.44}$$

for $i = 1, \ldots, n$.

With multiple dependent variables, it is more important than ever to use first integrals to simplify the Euler–Lagrange equations. If a dependent variable, $y_i(x)$, is missing from the integrand (4.33), the corresponding Euler–Lagrange equation simplifies to

$$\frac{\partial f}{\partial y_i'} \;=\; c\,. \tag{4.45}$$

Similarly, if the independent variable $x$ is missing from the integrand (4.33), the first term on the right side of

$$\frac{df}{dx} \;=\; \frac{\partial f}{\partial x} + \sum_{i=1}^{n}\left(\frac{\partial f}{\partial y_i}\, y_i' + \frac{\partial f}{\partial y_i'}\, y_i''\right) \tag{4.46}$$

disappears:

$$\frac{df}{dx} = \sum_{i=1}^{n} \left( \frac{\partial f}{\partial y_i} y_i' + \frac{\partial f}{\partial y_i'} y_i'' \right) . \tag{4.47}$$

Using our Euler–Lagrange equations, we may rewrite this last equation as

$$\begin{aligned} \frac{df}{dx} &= \sum_{i=1}^{n} \left[ \frac{d}{dx} \left( \frac{\partial f}{\partial y_i'} \right) y_i' + \frac{\partial f}{\partial y_i'} y_i'' \right] \\ &= \sum_{i=1}^{n} \frac{d}{dx} \left( \frac{\partial f}{\partial y_i'} y_i' \right) . \end{aligned} \tag{4.48}$$

It now follows that

$$\frac{d}{dx} \left( f - \sum_{i=1}^{n} \frac{\partial f}{\partial y_i'} y_i' \right) = 0 \tag{4.49}$$

and that

$$f - \sum_{i=1}^{n} \frac{\partial f}{\partial y_i'} y_i' = c \tag{4.50}$$

is a first integral.

## 4.4. Lagrangian mechanics

Problems with multiple dependent variables are especially common in mechanics, the science that studies the motion of material bodies. This science has a long history (Dugas, 1988). Newton laid the foundations for classical mechanics with his laws. Newtonian mechanics, which is built around vector quantities such as force, momentum, and acceleration, is quite general but can be unwieldy. Later scientists, most notably Joseph-Louis Lagrange and William Hamilton, reformulated mechanics so as to simplify matters. In this section, we will study Lagrangian mechanics (from a calculus of variations perspective). In a later section, we will study Hamiltonian mechanics.

Before proceeding, we must first consider some new notation. In classical mechanics, we are usually interested in the time evolution of dynamical systems. So time, $t$, is our independent variable. We also have dependent variables that specify the configuration of a mechanical system. These dependent variables need not be Cartesian coordinates. They can, for example, be angles or arc lengths. Mechanicians frequently denote their dependent variables, generically, as $q_i(t)$ and call them *generalized coordinates.*

For now, assume that we have chosen generalized coordinates that eliminate superfluous coordinates and explicit geometric constraints. (This is not always possible.) Thus, for a simple pendulum, which has one degree of freedom, we will use a single angle, rather than two Cartesian coordinates and a constraint on the length of the pendulum. For a double pendulum, with two degrees of freedom, we will use two angles rather than four Cartesian coordinates and two constraints. And, for a rigid body, with six degrees of freedom, we can use three coordinates to specify the center of the body and three Euler angles to specify the orientation of the body.

The generalized coordinates have time derivatives. We will denote these derivatives by $\dot{q}_i(t)$. Using a dot to signify a time derivative goes back to Newton and is still common in mechanics. Pulling everything together, our changes in notation are

$$x \rightarrow t\,, \quad y_i(x) \rightarrow q_i(t)\,, \quad y_i'(x) \rightarrow \dot{q}_i(t)\,. \tag{4.51}$$

For convenience, we will also use vector notation as needed.

The key advantage of Lagrangian mechanics is that we can often derive equations of motion using scalar quantities such as kinetic energy and potential energy. These scalar quantities are easier to work with than the vector quantities, force, momentum, and acceleration, that are part of Newtonian mechanics.

To keep things simple, consider a mechanical system in which all of the applied forces are derivable from a (time-dependent) scalar potential. Assume that we have a set of (independent) generalized coordinates that describes the configuration of this mechanical system; that we can express the kinetic energy $T$ of this system in terms of the generalized coordinates, their first derivatives, and time; and that we can express the potential energy $V$ of the system in terms of the generalized coordinates and time. Now form the *Lagrangian*

$$L(t, \mathbf{q}, \dot{\mathbf{q}}) \;\equiv\; T(t, \mathbf{q}, \dot{\mathbf{q}}) - V(t, \mathbf{q}) \tag{4.52}$$

and the time integral of the Lagrangian[1] along a path,

$$S[\mathbf{q}] \; = \; \int_{t_a}^{t_b} L(t, \mathbf{q}, \dot{\mathbf{q}}) \; dt\,. \tag{4.53}$$

Using these ingredients, Hamilton (1834, 1835) formulated the following important principle:

> ***Hamilton's principle:***
> The motion of a mechanical system from $\mathbf{q}(t_a) = \mathbf{q}_a$ to $\mathbf{q}(t_b) = \mathbf{q}_b$ is such that the first variation of the integral
>
> $$S[\mathbf{q}] \; = \; \int_{t_a}^{t_b} L(t, \mathbf{q}, \dot{\mathbf{q}}) \; dt \tag{4.54}$$
>
> is zero.

Imagine a mechanical system that moves from a given configuration at time $t_a$ to another specified configuration at time $t_b$. Imagine trajectories that take you between the two endpoints in the stipulated time. Hamilton's principle states that the true trajectory, $\hat{\mathbf{q}}(t)$, makes Hamilton's action *stationary*, $\delta S = 0$. In other words, the system follows the Euler–Lagrange equations

$$\frac{\partial L}{\partial q_i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) \; = \; 0 \tag{4.55}$$

for $i = 1, \ldots, n$.

Although a true trajectory makes Hamilton's action stationary, it need not make the action a minimum or a maximum. Phrased another way, observed trajectories are extremals (solutions of the Euler–Lagrange equations), but they do not necessarily yield extrema (maxima or minima) of Hamilton's action. At the same time, Hamilton's action is minimized in some cases. See Gelfand and Fomin (1963), Smith and Smith (1974), and Gray and Taylor (2007) for examples and discussion and see Joglekar and Tham (2011) for illustrations of

[1] There is no standard name for this integral. Many people refer to it, confusingly, as the *action* (confusingly because there is also an older integral, due to Maupertuis, named action). Others are more careful and say *Hamilton's action*. *Hamilton's first principle function* is another common name.

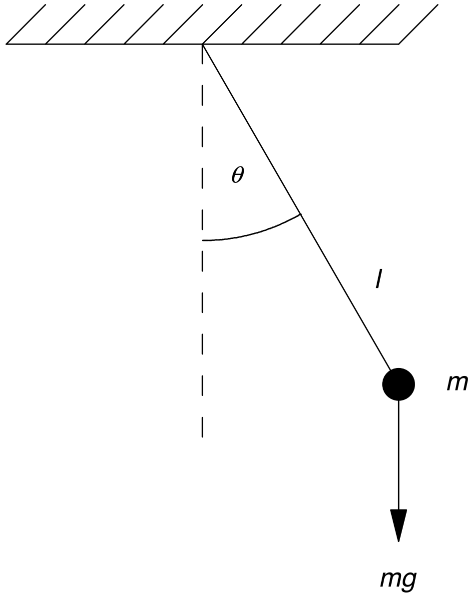



**Figure 4.4.** Simple pendulum

action landscapes. The situation is a bit like that in ordinary calculus, where setting a first derivative equal to zero might get you a saddle point rather than a minimum or a maximum. In this case, however, the saddle point has meaning.

**Example 4.1** (Simple pendulum)**.**

Consider a simple pendulum (see Figure 4.4) consisting of a mass $m$ at the end of a weightless string or rigid rod of length $l$ that swings back and forth in a vertical plane. We will use $\theta$, the angle that the string makes with the vertical, as our generalized coordinate. The kinetic energy of the mass is

$$T \;=\; \frac{1}{2}mv^2 \;=\; \frac{1}{2}m(l\dot{\theta})^2\,. \tag{4.56}$$

Its potential energy is

$$V(\theta) \;=\; mgl(1 - \cos\,\theta)\,. \tag{4.57}$$

The Lagrangian for this system is

$$L = T - V = \frac{1}{2}ml^2\dot{\theta}^2 - mgl\,(1 - \cos\,\theta) \tag{4.58}$$

and the Euler–Lagrange equation is

$$\frac{\partial L}{\partial \theta} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) = 0 \tag{4.59}$$

or

$$m\,l^2\,\ddot{\theta} = -mgl\,\sin\,\theta\,. \tag{4.60}$$

This is the equation of motion of the mass.

Since there is no explicit time dependence in our Lagrangian, we also have the first integral

$$L - \dot{\theta}\,\frac{\partial L}{\partial \dot{\theta}} = -E\,, \tag{4.61}$$

or

$$\frac{1}{2}ml^2\,\dot{\theta}^2 + mgl\,(1 - \cos\,\theta) = E\,, \tag{4.62}$$

where $E$ is the total energy of the system.

While the Lagrangian will produce the right equations of motion, it may not be the only integrand to do so in any particular instance. Consider, for example, a simple harmonic oscillator (with mass and spring constant one) with the equation of motion

$$\ddot{x} + x = 0\,. \tag{4.63}$$

Here, the displacement $x(t)$ is our generalized coordinate. You can derive this equation of motion using the Lagrangian

$$L(t, x, \dot{x}) = T(\dot{x}) - V(x) = \frac{1}{2}\dot{x}^2 - \frac{1}{2}x^2\,, \tag{4.64}$$

but you can also derive equation (4.63) using the Euler–Lagrange equation and the integrand

$$f(t, x, \dot{x}) = \frac{1}{x^2}(x\,\cos\,t - \dot{x}\,\sin\,t)\,\ln\left(\frac{x\,\cos\,t - \dot{x}\,\sin\,t}{x\,\sin\,t + \dot{x}\,\cos\,t}\right) \tag{4.65}$$

(Sarlet, 1981) or the integrand

$$f(t, x, \dot{x}) = \frac{\dot{x}}{x}\,\mathrm{atan}\left(\frac{\dot{x}}{x}\right) - \frac{1}{2}\,\ln\left[x^2\left(1 + \frac{\dot{x}^2}{x^2}\right)\right] \tag{4.66}$$

(Della Riccia, 1982).

Finding integrands that lead, by means of the calculus of variations, to a given set of differential equations is known as the *inverse problem of Lagrangian mechanics* or as the *inverse problem of the calculus of variations.* Reformulating differential equations as a variational problem can be useful: it can lead to new first integrals or to direct numerical methods for solving the equations.

It is easy to shed some light on the inverse problem for a system with one degree of freedom. Let us suppose that we wish to find the integral

$$J[x] \;=\; \int_{t_a}^{t_b} f(t, x, \dot{x})\; dt \tag{4.67}$$

that produces the equation of motion

$$\ddot{x} \;=\; g(t, x, \dot{x})\,. \tag{4.68}$$

The Euler–Lagrange equation for this system can be written in the ultradifferentiated form

$$\ddot{x}\,\frac{\partial^2 f}{\partial \dot{x}^2} + \dot{x}\,\frac{\partial^2 f}{\partial \dot{x}\partial x} + \frac{\partial^2 f}{\partial \dot{x}\partial t} - \frac{\partial f}{\partial x} \;=\; 0\,. \tag{4.69}$$

Let us take the partial derivative of this equation with respect to $\dot{x}$. The resulting equation is

$$\frac{\partial}{\partial \dot{x}}\left(g\,\frac{\partial^2 f}{\partial \dot{x}^2}\right) + \dot{x}\,\frac{\partial^3 f}{\partial \dot{x}^2\partial x} + \frac{\partial^3 f}{\partial \dot{x}^2\partial t} \;=\; 0\,, \tag{4.70}$$

where I have replaced $\ddot{x}$ with $g(t, x, \dot{x})$. This equation is actually a first-order partial differential equation in

$$M(t, x, \dot{x}) \;\equiv\; \frac{\partial^2 f}{\partial \dot{x}^2} \tag{4.71}$$

that can be written

$$\frac{\partial}{\partial \dot{x}}(gM) + \dot{x}\,\frac{\partial M}{\partial x} + \frac{\partial M}{\partial t} \;=\; 0\,. \tag{4.72}$$

This equation is sometimes known as *Jacobi's equation of the last multiplier,* with $M(t, x, \dot{x})$ as the last multiplier.

In many cases, particular solutions of equation (4.72) are easy to find, even if the general solution is hard to find. Once a nontrivial solution of equation (4.72) has been found, definition (4.71) implies that we can integrate $M(t,x,\dot{x})$ twice with respect $\dot{x}$ to obtain appropriate integrands. The resulting constants of integration are not arbitrary but must be chosen so that $f(t,x,\dot{x})$ still satisfies our Euler–Lagrange equation. See Vujanovic and Jones (1989) for examples and for further information regarding the inverse problem of the calculus of variations.

We still haven't attacked a real mechanics problem with multiple dependent variables. Let's try one now.

## 4.5. Case study: The spherical pendulum

Let us consider a pendulum, consisting of a mass $m$ at the end of a weightless string or rigid rod of length $l$, with two degrees of freedom (see Figure 4.5). Because of the string, the mass is constrained to move on the surface of a sphere. We will let $\theta$, the angle from the downward vertical, and $\phi$, the azimuthal angle, be our generalized coordinates.

In this coordinate system, the scale factors are

$$h_\theta = l\,, \quad h_\phi = l\,\sin\theta \tag{4.73}$$

and the kinetic energy is

$$\begin{aligned} T &= \frac{1}{2}\,mv^2 = \frac{1}{2}m[(h_\theta\dot{\theta})^2 + (h_\phi\dot{\phi})^2] \\ &= \frac{1}{2}\,m\,l^2\,(\dot{\theta}^2 + \dot{\phi}^2\,\sin^2\theta)\,. \end{aligned} \tag{4.74}$$

The potential energy, in turn, is the same as that for the plane pendulum,

$$V(\theta) = mgl\,(1-\cos\theta)\,. \tag{4.75}$$

The Lagrangian is thus

$$L = \frac{1}{2}\,ml^2(\dot{\theta}^2 + \dot{\phi}^2\,\sin^2\theta) - mgl\,(1-\cos\theta)\,. \tag{4.76}$$

There are two equations of motion,

$$\frac{\partial L}{\partial\theta} - \frac{d}{dt}\left(\frac{\partial L}{\partial\dot{\theta}}\right) = 0\,, \quad \frac{\partial L}{\partial\phi} - \frac{d}{dt}\left(\frac{\partial L}{\partial\dot{\phi}}\right) = 0\,, \tag{4.77}$$

one for each degree of freedom.

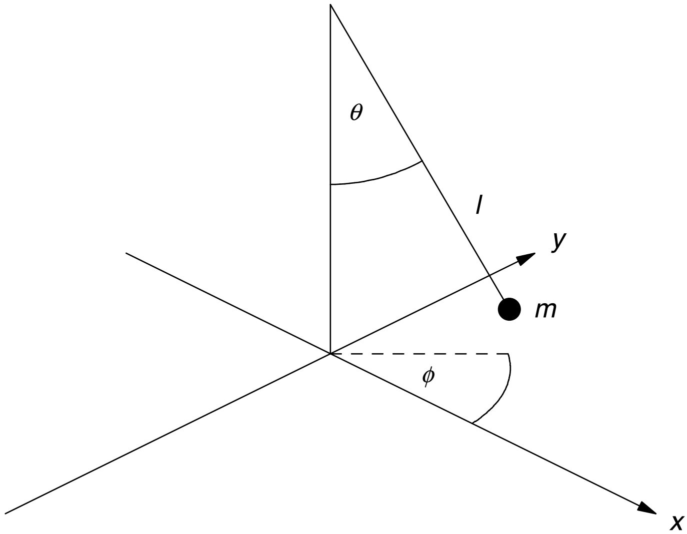



**Figure 4.5.** Spherical pendulum

We may look for the usual first integrals. If a coordinate $q_i$ fails to appear in the Lagrangian, it is ignorable and

$$\frac{\partial L}{\partial \dot{q}_i} \;=\; p_{q_i}\,, \tag{4.78}$$

$p_{q_i}$ a constant. If the Lagrangian is autonomous, with no explicit dependence on $t$,

$$L - \sum_{i=1}^{n} \dot{q}_i \, \frac{\partial L}{\partial \dot{q}_i} \;=\; -E\,, \tag{4.79}$$

with $E$ a constant.

For the spherical pendulum, $\phi$, which does not appear in the Lagrangian, is an ignorable coordinate. It follows that

$$\frac{\partial L}{\partial \dot{\phi}} \;=\; ml^2\, \dot{\phi}\, \sin^2\theta \;=\; p_\phi\,. \tag{4.80}$$

The Lagrangian is also autonomous so that

$$L - \dot{\theta}\, \frac{\partial L}{\partial \dot{\theta}} + \dot{\phi}\, \frac{\partial L}{\partial \dot{\phi}} \;=\; -E\,. \tag{4.81}$$

Thus

$$\frac{1}{2}ml^2(\dot\theta^2 + \dot\phi^2 \sin^2\theta) + mgl\,(1 - \cos\theta) \;=\; E\,, \tag{4.82}$$

where $E$ is the total mechanical energy of the system.

We may eliminate $\dot\phi$ from this last expression using the first integral for our ignorable variable,

$$\frac{1}{2}ml^2\dot\theta^2 + \frac{p_\phi^2}{2ml^2\sin^2\theta} + mgl\,(1 - \cos\theta) \;=\; E\,. \tag{4.83}$$

This last equation only contains the variables $\theta$ and $\dot\theta$ and can be reduced to quadratures. Indeed, one can write an *effective* potential for the motion,

$$U(\theta) \;=\; mgl\,(1 - \cos\theta) + \frac{p_\phi^2}{2ml^2\sin^2\theta}\,, \tag{4.84}$$

so that

$$\frac{1}{2}ml^2\,\dot\theta^2 \;=\; E - U(\theta)\,. \tag{4.85}$$

Since the left-hand side of this equation is nonnegative, the motion is confined to those values of $\theta$ such that

$$U(\theta) \;\leq\; E\,. \tag{4.86}$$

To see what is going on (or around), we may plot the effective potential $U(\theta)$ as a function of $\theta$. For $p_\phi = 0$, the potential curve is that of a simple pendulum. It has a minimum at $\theta = 0$ and a maximum at $\theta = \pi$. For $p_\phi \neq 0$, the effective potential has a minimum at an angle, $\theta^*$, that is the root of

$$\frac{dU}{d\theta} \;=\; mgl\,\sin\theta - \frac{p_\phi^2\cos\theta}{ml^2\sin^3\theta} \;=\; 0\,. \tag{4.87}$$

It is clear, from this equation, that $0 < \theta^* < \pi/2$ and that $\theta^*$ approaches $\pi/2$ as $p_\phi$ approaches infinity. For $E > U(\theta^*)$, $\theta$ oscillates between a minimum and a maximum value while the mass swings about the vertical or $z$-axis.

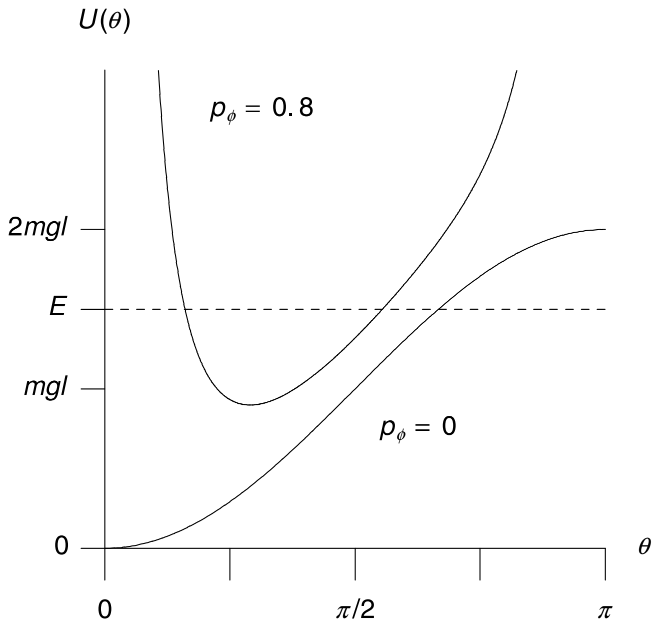



**Figure 4.6.** Effective potential

## 4.6. Hamiltonian mechanics

Now is a good time to introduce some additional ideas from mechanics that are useful throughout the calculus of variations. In general, the Euler–Lagrange equations are a system of $n$ second-order differential equations. It is often convenient to rewrite these equations as a system of $2n$ first-order differential equations. There are many ways to do this, but there is one particularly useful set of transformations, closely related to our standard first integrals, that gives rise to the Hamiltonian formulation of mechanics.

Let us introduce the canonical momenta

$$p_i \;=\; p_i(t, \mathbf{q}, \dot{\mathbf{q}}) \;\equiv\; \frac{\partial}{\partial \dot{q}_i} L(t, \mathbf{q}, \dot{\mathbf{q}}) \qquad (4.88)$$

for $i = 1, \ldots, n$. In what follows, we will refer to both the variables $p_i$ and the functions $p_i(t, \mathbf{q}, \dot{\mathbf{q}})$. Pay careful attention.

Ideally, we can now rewrite the time derivatives of the generalized coordinates as functions of time $t$, the generalized coordinates $\mathbf{q}$, and the canonical momenta $\mathbf{p}$,

$$\dot{q}_i = \dot{q}_i(t, \mathbf{q}, \mathbf{p}) . \tag{4.89}$$

By the implicit function theorem, this is possible (at least locally) if the matrix

$$\left[ \frac{\partial p_i}{\partial q_j} \right] = \left[ \frac{\partial^2 L(t, \mathbf{q}, \dot{\mathbf{q}})}{\partial \dot{q}_i \partial \dot{q}_j} \right] , \tag{4.90}$$

with row index $i = 1, \ldots, n$ and column index $j = 1, \ldots, n$, is nonsingular (has nonzero determinant).

Let us now create the *Hamiltonian,* by means of the *Legendre transformation*

$$\begin{aligned} H(t, \mathbf{q}, \mathbf{p}) &\equiv \mathbf{p} \cdot \dot{\mathbf{q}}(t, \mathbf{q}, \mathbf{p}) - L(t, \mathbf{q}, \dot{\mathbf{q}}(t, \mathbf{q}, \mathbf{p})) \qquad (4.91) \\ &= \left[ \sum_{i=1}^{n} p_i \dot{q}_i(t, \mathbf{q}, \mathbf{p}) \right] - L(t, \mathbf{q}, \dot{\mathbf{q}}(t, \mathbf{q}, \mathbf{p})) . \end{aligned}$$

The Legendre transformation is *involutive* (it is its own inverse), and it is easy to see that

$$\begin{aligned} L(t, \mathbf{q}, \dot{\mathbf{q}}) &= \mathbf{p}(t, \mathbf{q}, \dot{\mathbf{q}}) \cdot \dot{\mathbf{q}} - H(t, \mathbf{q}, \mathbf{p}(t, \mathbf{q}, \dot{\mathbf{q}})) \qquad (4.92) \\ &= \left[ \sum_{i=1}^{n} p_i(t, \mathbf{q}, \dot{\mathbf{q}})\, \dot{q}_i \right] - H(t, \mathbf{q}, \mathbf{p}(t, \mathbf{q}, \dot{\mathbf{q}})) . \end{aligned}$$

See Zia et al. (2009) for more on the Legendre transformation.

What are the equations of motion in terms of the Hamiltonian? Let us begin by taking the partial derivative

$$\begin{aligned} \frac{\partial H}{\partial p_i} &= \dot{q}_i(t, \mathbf{q}, \mathbf{p}) \qquad (4.93) \\ &\quad + \sum_{j=1}^{n} \left[ p_j \frac{\partial}{\partial p_i} \dot{q}_j(t, \mathbf{q}, \mathbf{p}) - \frac{\partial L}{\partial q_j} \frac{\partial q_j}{\partial p_i} - \frac{\partial L}{\partial \dot{q}_j} \frac{\partial}{\partial p_i} \dot{q}_j(t, \mathbf{q}, \mathbf{p}) \right] \end{aligned}$$

so that

$$\frac{\partial H}{\partial p_i} = \dot{q}_i(t, \mathbf{q}, \mathbf{p}) + \sum_{j=1}^{n} \left[ \left( p_j - \frac{\partial L}{\partial \dot{q}_j} \right) \frac{\partial}{\partial p_i} \dot{q}_j(t, \mathbf{q}, \mathbf{p}) - \frac{\partial L}{\partial q_j} \frac{\partial q_j}{\partial p_i} \right] . \tag{4.94}$$

The first term in the square brackets vanishes because of our definition of the canonical momenta while the second term in the square brackets vanishes because the $q_i$ and the $p_i$ are independent variables in the Hamiltonian framework.

Let us now look at the partial derivative

$$\frac{\partial H}{\partial q_i} = \sum_{j=1}^{n} \left[ p_j \frac{\partial}{\partial q_i} \dot{q}_j(t, \mathbf{q}, \dot{\mathbf{q}}) + \frac{\partial p_j}{\partial q_i} \dot{q}_j(t, \mathbf{q}, \dot{\mathbf{q}}) \right] - \frac{\partial L}{\partial q_i} - \sum_{j=1}^{n} \frac{\partial L}{\partial \dot{q}_j} \frac{\partial}{\partial q_i} \dot{q}_j(t, \mathbf{q}, \dot{\mathbf{q}}) . \tag{4.95}$$

The first and fourth terms on the right-hand side cancel because of our definition of the canonical momenta while the second term vanishes because of the independence of the $q_i$ and the $p_i$. Finally, the $\partial L/\partial q_i$ are equal to $\dot{p}_i$ by the Euler–Lagrange equations.

Combining these results, we obtain

$$\dot{q}_i = \frac{\partial H}{\partial p_i} , \quad \dot{p}_i = -\frac{\partial H}{\partial q_i} , \tag{4.96}$$

for $i = 1, \ldots, n$. These equations are the famous *canonical* or *Hamiltonian* equations of motion. These $2n$ canonical equations are equivalent to the $n$ Euler–Lagrange equations.

**Example 4.2** (Harmonic oscillator)**.**

Consider a mass $m$ at the end of a horizontal spring with spring constant $k$. The kinetic energy

$$T(\dot{x}) = \frac{1}{2} m \dot{x}^2 \tag{4.97}$$

and the potential energy

$$V(x) = \frac{1}{2} k x^2 \tag{4.98}$$

lead to the Lagrangian

$$L(x,\dot{x}) \;=\; \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2\,. \tag{4.99}$$

The Euler–Lagrange equation is simply

$$\frac{\partial L}{\partial x} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) \;=\; 0 \tag{4.100}$$

or

$$m\ddot{x} \;=\; -kx\,. \tag{4.101}$$

The canonical momentum for the harmonic oscillator is

$$p \;=\; p(\dot{x}) \;=\; \frac{\partial L}{\partial \dot{x}} \;=\; m\dot{x} \tag{4.102}$$

and we may write the velocity $\dot{x}$, in terms of the momentum, as

$$\dot{x} \;=\; \dot{x}(p) \;=\; \frac{p}{m}\,. \tag{4.103}$$

Using the Legendre transformation, the Hamiltonian is

$$H(x,p) \;=\; p\,\dot{x}(p) - L(x,\dot{x}(p)) \;=\; \frac{p^2}{m} - \left(\frac{p^2}{2m} - \frac{1}{2}\,kx^2\right) \tag{4.104}$$

or

$$H(x,p) \;=\; \frac{p^2}{2m} + \frac{1}{2}kx^2\,. \tag{4.105}$$

For this Hamiltonian, the canonical equations imply that

$$\dot{x} \;=\; \frac{\partial H}{\partial p} \;=\; \frac{p}{m}\,, \quad \dot{p} \;=\; -\frac{\partial H}{\partial x} \;=\; -kx\,. \tag{4.106}$$

Our interest is not in the canonical equations per se, but in the fact that expressions identical to or similar to the canonical momenta and the Hamiltonian frequently occur in the calculus of variations. Consider, for example, our standard first integrals. If a cyclic or ignorable coordinate, $q_i$, is missing from the Hamiltonian, then, by the second canonical equation,

$$\dot{p}_i \;=\; -\frac{\partial H}{\partial q_i} \;=\; 0 \tag{4.107}$$

so that $p_i$ is constant. This is the canonical analog of first integral (4.45).

Next, consider the case where our Lagrangian has no explicit time dependence. Then, by equations (4.88) and (4.91), our canonical momenta and Hamiltonian have no explicit time dependence. As a result,

$$\frac{dH}{dt} = \sum_{i=1}^{n} \left( \frac{\partial H}{\partial q_i} \frac{dq_i}{dt} + \frac{\partial H}{\partial p_i} \frac{dp_i}{dt} \right) + \frac{\partial H}{\partial t} \tag{4.108}$$

simplifies to

$$\frac{dH}{dt} = \sum_{i=1}^{n} \left( \frac{\partial H}{\partial q_i} \frac{dq_i}{dt} + \frac{\partial H}{\partial p_i} \frac{dp_i}{dt} \right) . \tag{4.109}$$

Along our extremals, however, the canonical equations are satisfied. So, if we substitute $\dot{q}_i$ and $\dot{p}_i$ from equation (4.96), $dH/dt = 0$. Thus, if there is no explicit time dependence, our Hamiltonian is a constant of motion. This is the canonical analog of conservation law (4.50).

The Hamiltonian, strictly speaking, is a function of time, the generalized coordinates, and the canonical momenta. The closely related function

$$h(t, \mathbf{q}, \dot{\mathbf{q}}) = \left( \sum_{i=1}^{n} \frac{\partial L}{\partial \dot{q}_i} \dot{q}_i \right) - L(t, \mathbf{q}, \dot{\mathbf{q}}) , \tag{4.110}$$

which appears frequently in Lagrangian mechanics, has different arguments but is identical in value to the Hamiltonian. It is sometimes called the *energy function* (Goldstein, 1980) to distinguish it from the Hamiltonian. Constancy of the energy function is sometimes called the *Jacobi conservation law* or the *Jacobi integral* (Goldstein, 1980; Vujanovic and Jones, 1989; Vujanovic and Atanackovic, 2004). In some sense, the energy function is merely a surrogate for the Hamiltonian. The Hamiltonian, the canonical momenta, and their stand-ins will appear frequently throughout this course.

You will find more opportunities to work with mechanics problems with multiple dependent variables in the exercises at the end of the chapter. Don't, however, think that mechanics is the only place where multiple dependent variables arise. Indeed, let us briefly discuss some optics.

## 4.7. Ray optics

Ray optics is the simplest theory of light. In this theory, light travels as rays through an optical medium. An optical medium is characterized by its *refractive index,* $n \geq 1$. This index is the ratio of the speed of light in free space, $c$, to that in the medium, $v$. The time $T$ taken by light to travel a distance $d$ is thus

$$T = \frac{d}{v} = \frac{nd}{c}. \tag{4.111}$$

This time is proportional to $nd$, which is known as the *optical path length.*

In an inhomogeneous medium, the refractive index is a function of position,

$$n = n(x,y,z). \tag{4.112}$$

The optical path length in an inhomogeneous medium is thus given by

$$cT = \int n(x,y,z)\, ds = \int n(x,y,z)\sqrt{dx^2+dy^2+dz^2} \tag{4.113}$$

where the integral is taken over the given path or curve. Fermat's principle states that an optical ray, traveling between two points, follows a path that causes the travel time and the optical path length to be stationary relative to neighboring paths.

If the optical medium has a preferred axis (see the next example), you may use the corresponding spatial coordinate as your independent variable. If you don't want to single out one of the spatial coordinates, you may instead use time (or, more generally, some arbitrary stepping parameter) as your independent variable. The optical path length is then

$$cT = \int n(x,y,z)\frac{ds}{dt}\, dt = \int n(x,y,z)\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}\, dt. \tag{4.114}$$

Since time does not explicitly appear in our integrand, you may be tempted to write the corresponding first integral. Unfortunately, our integrand is positive-homogeneous of degree one in its derivatives and for such problems the first integral is always a tautology. (We will discuss homogeneous problems in more detail in Chapter 8.) Fortunately, the Euler–Lagrange equations,

$$\dot{s}\,\frac{\partial n}{\partial x} - \frac{d}{dt}\left(\frac{n\dot{x}}{\dot{s}}\right) = 0\,, \quad \dot{s}\,\frac{\partial n}{\partial y} - \frac{d}{dt}\left(\frac{n\dot{y}}{\dot{s}}\right) = 0\,, \tag{4.115}$$
$$\dot{s}\,\frac{\partial n}{\partial z} - \frac{d}{dt}\left(\frac{n\dot{z}}{\dot{s}}\right) = 0$$

or

$$\frac{\partial n}{\partial x} - \frac{d}{ds}\left(n\frac{dx}{ds}\right) = 0\,, \quad \frac{\partial n}{\partial y} - \frac{d}{ds}\left(n\frac{dy}{ds}\right) = 0\,, \tag{4.116}$$
$$\frac{\partial n}{\partial z} - \frac{d}{ds}\left(n\frac{dz}{ds}\right) = 0\,,$$

are more helpful.

Physicists often write the last system of equations in the vector form

$$\frac{d}{ds}\left(n\frac{d\mathbf{r}}{ds}\right) = \nabla n\,. \tag{4.117}$$

This equation, known as the *ray equation,* is meant to be an optical analog to Newton's law (Evans and Rosenquist, 1986), with $\nabla n$ playing the role of force and the term in parentheses playing the role of momentum. In a homogeneous medium, with the $n$ a constant, the ray equation reduces to $\mathbf{r}''(s) = 0$. In other words, light rays are straight lines in a homogeneous medium. Life is more interesting when the refractive index varies.

**Example 4.3** (Fiber optics)**.**

An optical fiber is a cylindrical waveguide made of a low-loss material such as silica glass. Light is guided through a central core that is embedded in an outer cladding. This cladding has lower refractive index than the core. Light rays that graze the core-cladding boundary at a shallow angle undergo total internal reflection and are guided through the core.

Conventional fibers have constant refractive indices in the core and the cladding and are known as *step-index fibers.* There are also *graded-index fibers* that have a refractive index that decreases continuously from its center of the fiber. What is the path of light in a graded-index fiber?

Let's attack this problem using cylindrical coordinates, with $z$ as the coordinate along the axis of the fiber. According to Fermat's principle, the path connecting two arbitrary points, $(r_1, \theta_1, z_1)$ and $(r_2, \theta_2, z_2)$, makes the optical path length

$$cT = \int n(r)\sqrt{dr^2 + (r\,d\theta)^2 + dz^2} \tag{4.118}$$

stationary. If we choose $z$ as the independent variable, our path length becomes

$$cT = \int_{z_1}^{z_2} n(r)\sqrt{(r')^2 + (r\theta')^2 + 1}\,dz \tag{4.119}$$

where

$$r' = \frac{dr}{dz} \quad \text{and} \quad \theta' = \frac{d\theta}{dz}\,. \tag{4.120}$$

The functions $r(z)$ and $\theta(z)$ that describe the path with the fiber satisfy the two Euler–Lagrange equations

$$\frac{\partial f}{\partial r} - \frac{d}{dz}\left(\frac{\partial f}{\partial r'}\right) = 0\,, \quad \frac{\partial f}{\partial \theta} - \frac{d}{dz}\left(\frac{\partial f}{\partial \theta'}\right) = 0\,, \tag{4.121}$$

where

$$f = n(r)\sqrt{(r')^2 + (r\theta')^2 + 1}\,. \tag{4.122}$$

Since $\theta$ is a cycle or ignorable variable, we obtain, as one first integral,

$$\frac{\partial f}{\partial \theta'} = \frac{n(r)\,r^2\,\theta'}{\sqrt{(r')^2 + (r\theta')^2 + 1}} = \alpha\,. \tag{4.123}$$

In addition, since $z$ does not appear explicitly in our integrand,

$$f - r'\frac{\partial f}{\partial r'} - \theta'\frac{\partial f}{\partial \theta'} = \frac{n(r)}{\sqrt{(r')^2 + (r\theta')^2 + 1}} = \beta \tag{4.124}$$

is another first integral. These two coupled, first-order, ordinary differential equations enable us to determine $r(z)$ and $\theta(z)$ and to determine the path of light through a graded-index optic fiber.

If, in first integral (4.123), $\theta'(0) = 0$, then $\alpha = 0$ and $\theta'(z) = 0$ for all $z$. Rays then remain within a constant $\theta$ plane that passes through the axis of symmetry; these rays are known as *meridional* rays. For $\theta' = 0$, equation (4.124) then simplifies significantly.

## 4.8. Double integrals

Finally, consider the problem of minimizing a functional of the form

$$J[u] \;=\; \iint_A f(x,y,u,u_x,u_y)\;dx\,dy\,, \tag{4.125}$$

where $u = u(x,y)$ is a function of both $x$ and $y$, subject to the boundary condition

$$u(x,y) \;=\; u_0(x,y) \tag{4.126}$$

for all points $(x,y)$ on the boundary curve $\partial A$.

Let us now suppose that $u = \hat{u}(x,y)$ solves this problem and that $\epsilon\eta(x,y)$ is a weak variation,

$$u(x,y) \;=\; \hat{u}(x,y) \;+\; \epsilon\,\eta(x,y)\,, \tag{4.127}$$

with

$$\eta(x,y) \;=\; 0 \tag{4.128}$$

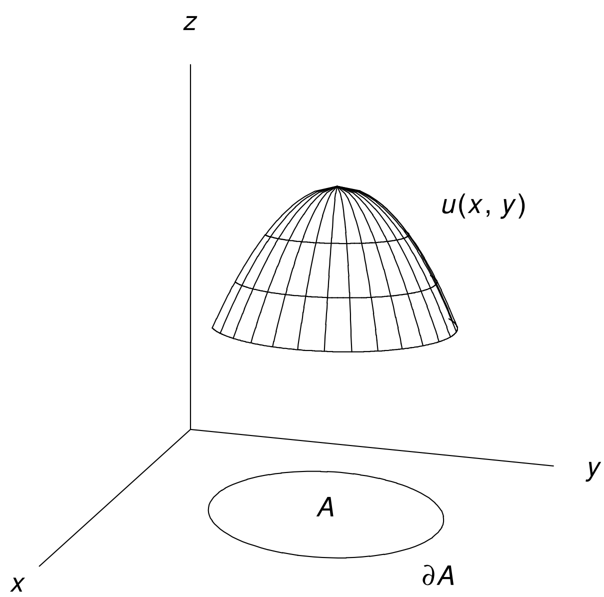



**Figure 4.7.** Function of two variables

at all points of $\partial A$. The total variation is now

$$\Delta J[u] \;=\; J[u] - J[\hat{u}] \;=\; J[\hat{u} + \epsilon\,\eta] - J[\hat{u}] \tag{4.129}$$

or

$$\begin{aligned} \Delta J[u] \;=\; & \iint_A f(x, y, \hat{u} + \epsilon\eta, \hat{u}_x + \epsilon\eta_x, \hat{u}_y + \epsilon\eta_y)\; dx\, dy \\ & - \iint_A f(x, y, \hat{u}, \hat{u}_x, \hat{u}_y)\; dx\, dy\,. \end{aligned} \tag{4.130}$$

If we expand this total variation in a Taylor series in $\epsilon$ in the usual way, we obtain

$$\Delta J \;=\; \delta J + \frac{1}{2}\,\delta^2 J + \cdots \tag{4.131}$$

where

$$\delta J \;=\; \epsilon \iint_A \left( \frac{\partial f}{\partial u}\,\eta + \frac{\partial f}{\partial u_x}\,\eta_x + \frac{\partial f}{\partial u_y}\,\eta_y \right)\; dx\, dy\,. \tag{4.132}$$

Note that

$$\frac{\partial f}{\partial u_x}\,\eta_x \;=\; \frac{\partial}{\partial x}\left(\eta\,\frac{\partial f}{\partial u_x}\right) - \eta\,\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial u_x}\right), \tag{4.133}$$

$$\frac{\partial f}{\partial u_y}\,\eta_y \;=\; \frac{\partial}{\partial y}\left(\eta\,\frac{\partial f}{\partial u_y}\right) - \eta\,\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial u_y}\right). \tag{4.134}$$

The first variation may, in other words, be rewritten as

$$\begin{aligned} \delta J \;=\; & \epsilon \iint_A \left[ \frac{\partial f}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial u_x}\right) - \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial u_y}\right) \right] \eta\; dx\, dy \\ & + \epsilon \iint_A \left[ \frac{\partial}{\partial x}\left(\eta\,\frac{\partial f}{\partial u_x}\right) + \frac{\partial}{\partial y}\left(\eta\,\frac{\partial f}{\partial u_y}\right) \right]\; dx\, dy\,. \end{aligned} \tag{4.135}$$

The second integral in the last equation can be transformed to a line integral over $\partial A$ using Green's theorem,

$$\iint_A \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right)\; dx\, dy \;=\; \int_{\partial A} P\; dx \,+\, Q\; dy\,. \tag{4.136}$$

It now follows that

$$\delta J = \epsilon \iint\limits_A \left[ \frac{\partial f}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial u_x}\right) - \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial u_y}\right) \right] \eta \, dx\, dy \qquad (4.137)$$

$$+ \epsilon \int\limits_{\partial A} \left(-\eta \frac{\partial f}{\partial u_y}\right) dx + \left(\eta \frac{\partial f}{\partial u_x}\right) dy\,.$$

Since $\eta(x, y)$ vanishes on $\partial A$ and since $\eta(x, y)$ is otherwise arbitrary, the Euler–Lagrange equation is now the partial differential equation

$$\frac{\partial f}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial u_x}\right) - \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial u_y}\right) = 0\,. \qquad (4.138)$$

**Example 4.4** (Plateau's problem)**.**

Consider the surface area integral

$$J[u] = \iint\limits_A \sqrt{1 + u_x^2 + u_y^2}\; dx\, dy\,. \qquad (4.139)$$

This leads to the partial differential equation

$$\frac{\partial}{\partial x}\left(\frac{u_x}{\sqrt{1 + u_x^2 + u_y^2}}\right) + \frac{\partial}{\partial y}\left(\frac{u_y}{\sqrt{1 + u_x^2 + u_y^2}}\right) = 0\,, \qquad (4.140)$$

which simplifies to

$$\frac{(1 + u_y^2)\, u_{xx} - 2\, u_x\, u_y\, u_{xy} + (1 + u_x^2)\, u_{yy}}{(1 + u_x^2 + u_y^2)^{3/2}} = 0 \qquad (4.141)$$

or

$$(1 + u_y^2)\, u_{xx} - 2\, u_x\, u_y\, u_{xy} + (1 + u_x^2)\, u_{yy} = 0\,. \qquad (4.142)$$

This is the minimal surface equation that we considered in Chapter 1. For small deflections, this equation may be approximated by Laplace's equation.

In Chapter 1, we also saw that the minimal surface equation can be given a geometric interpretation. At each point $P$ of our surface, choose a vector normal to the surface, cut the surface with normal planes (that contain the normal vector but that differ in orientation), and obtain a series of plane curves. For each plane curve, determine the curvature at $P$. Find the minimum and maximum curvatures,

$\kappa_{\min}$ and $\kappa_{\max}$, from amongst all the plane curves passing through $P$. These are your *principal curvatures.*

Instructors of differential geometry courses often prove that the mean curvature,

$$H = \frac{\kappa_{\max} + \kappa_{\min}}{2}, \tag{4.143}$$

the average of the maximum and minimum normal curvatures, is just

$$H = \frac{(1 + u_y^2)\, u_{xx} - 2\, u_x\, u_y\, u_{xy} + (1 + u_x^2)\, u_{yy}}{2\,(1 + u_x^2 + u_y^2)^{3/2}}. \tag{4.144}$$

Hence, the mean curvature is zero at every point of a minimizing surface: at each point, the surface is either flat or looks like a saddle. Surfaces with zero mean curvature are traditionally called minimal surfaces, whether they minimize area or not. Those that do minimize area are then said to be *stable* minimal surfaces.

A new spatial variable can also help one construct a Lagrangian for a problem with a continuum of generalized coordinates. One then talks of the *Lagrangian density* with respect to this new spatial variable.

**Example 4.5** (D'Alembert's wave equation)**.**

Consider a string of infinitesimal thickness and line density $\rho$ that is stretched with constant tensile force (tension) $\tau$ between $x = 0$ and $x = l$. Let $u(t, x)$ be the vertical displacement of this string.

The kinetic energy of the string is

$$T = \frac{1}{2} \int_0^l \rho\, u_t^2 \; dx\,. \tag{4.145}$$

Let us now consider a small element of the string of length $\Delta x$. As we displace this element, we stretch it by an amount

$$\Delta s - \Delta x = \left[ \sqrt{1 + \left( \frac{\partial u}{\partial x} \right)^2} - 1 \right] \Delta x\,, \tag{4.146}$$

where $\Delta s$ is the arc length of the stretched string. Using a binomial series,

$$\Delta s - \Delta x \approx \frac{1}{2}\left(\frac{\partial u}{\partial x}\right)^2 \Delta x \,. \tag{4.147}$$

The potential energy of the string is thus

$$V = \frac{\tau}{2}\int_0^l u_x^2 \, dx \,. \tag{4.148}$$

The above kinetic and potential energies suggest the Lagrangian density

$$\mathcal{L}(t, x, u, u_t, u_x) = \frac{1}{2}\rho\, u_t^2 - \frac{1}{2}\tau\, u_x^2 \,, \tag{4.149}$$

the Lagrangian

$$L = \int_0^l \mathcal{L}(t, x, u, u_t, u_x) \, dx \,, \tag{4.150}$$

and the functional

$$\int_{t_a}^{t_b}\int_0^l \mathcal{L}(t, x, u, u_t, u_x) \, dx\, dt \,. \tag{4.151}$$

The Euler–Lagrange equation for this double integral,

$$\frac{\partial \mathcal{L}}{\partial u} - \frac{\partial}{\partial t}\frac{\partial \mathcal{L}}{\partial u_t} - \frac{\partial}{\partial x}\frac{\partial \mathcal{L}}{\partial u_x} = 0 \,, \tag{4.152}$$

reduces to the wave equation

$$\rho\,\frac{\partial^2 u}{\partial t^2} = \tau\,\frac{\partial^2 u}{\partial x^2} \,. \tag{4.153}$$

Many linear and nonlinear wave equations are derived using variational principles. The calculus of variations for multiple integrals is an active and ongoing area of research.

## 4.9. Recommended reading

For further discussion of problems with second-order derivatives, two or more dependent variables, and/or double integrals, see the classic and encyclopedic book by Forsyth (1927).

Problems with second-order derivatives commonly appear in solid mechanics. For more on variational methods in solid mechanics, see the books by Haichang (1984), Wan (1995), Reddy (2002), Wallerstein (2002), and Rao (2007).

The literature on variational methods in analytical (Lagrangian and Hamiltonian) mechanics is substantial. One of the most enjoyable books on the topic is due to Lanczos (1974). This book has been reprinted as an affordable Dover paperback. Other useful books include Yourgrau and Mandelstam (1968), Tabarrok and Rimrott (1994), Vujanovic and Atanackovic (2004), and Basdevant (2007). Nakane and Fraser (2002) provide a detailed analysis of Hamilton's 1834 and 1835 papers. See Mercier (1959) and Moiseiwitsch (1966) for more on Lagrangian densities.

Lemons (1997) provides a gentle introduction to the calculus of variations in ray optics while Stavroudis (1972), Born and Wolf (1999), Lakshminarayanan et al. (2002), and Stavroudis (2006) consider this topic in greater detail. Marchand (1978) provides an overview of graded-index optics; Okoshi (1982) covers optical fibers.

Fermat's principle is used in ray acoustics, seismology, and other fields besides ray optics. Slawinski (2010) shows how seismic problems in elastic continua can be formulated and solved using the calculus of variations. Kimball and Story (1998) discuss how Fermat's principle can be used in sailing.

## 4.10. Exercises

**4.10.1. A degenerate case.** Prove (Forsyth, 1927) that, for

$$f(x,y,y',y'') \;=\; M(x,y,y') + N(x,y,y')\,y''\,, \tag{4.154}$$

equation (4.10) reduces to a differential equation of, at most, order two.

**4.10.2. An identity.** Show (Forsyth, 1927) that, for equation (4.10) to be an identity rather than a differential equation, we must have

$$f(x, y, y', y'') = \frac{\partial S}{\partial x} + \frac{\partial S}{\partial y} y' + \frac{\partial S}{\partial y'} y'', \tag{4.155}$$

where $S$ is any function of $x$, $y$, and $y'$.

**4.10.3. First integrals.** Prove (van Brunt, 2004) that

(a) if $f(x, y, y', y'')$ does not depend on $y$ explicitly, equation (4.10) has the first integral

$$\frac{d}{dx}\left(\frac{\partial f}{\partial y''}\right) - \frac{\partial f}{\partial y'} = c, \tag{4.156}$$

$c$ a constant, and that

(b) if $f(x, y, y', y'')$ does not depend on $x$ explicitly, equation (4.10) has the first integral

$$y'' \frac{\partial f}{\partial y''} - y' \left[\frac{d}{dx}\left(\frac{\partial f}{\partial y''}\right) - \frac{\partial f}{\partial y'}\right] - f = c, \tag{4.157}$$

with $c$ again a constant.

**4.10.4. Derivation of the Euler–Poisson equation.** Prove for an integral

$$J = \int_a^b f(x, y(x), y'(x), \ldots, y^{(n)}(x))\, dx \tag{4.158}$$

that contains derivatives of order $n$ that $\hat{y}(x)$ must satisfy the Euler–Poisson equation

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) + \cdots + (-1)^n \frac{d^n}{dx^n}\left[\frac{\partial f}{\partial y^{(n)}}\right] = 0. \tag{4.159}$$

**4.10.5. Curvature in the plane.** The curvature $\kappa$ of a plane curve $y(x)$ is the instantaneous rate of change of the slope angle $\phi$, from the $x$-axis to the tangent to the curve, with respect to arc length $s$. Starting with this definition, show that

$$\kappa(x) = \frac{y''(x)}{(1 + y'^2)^{3/2}}. \tag{4.160}$$

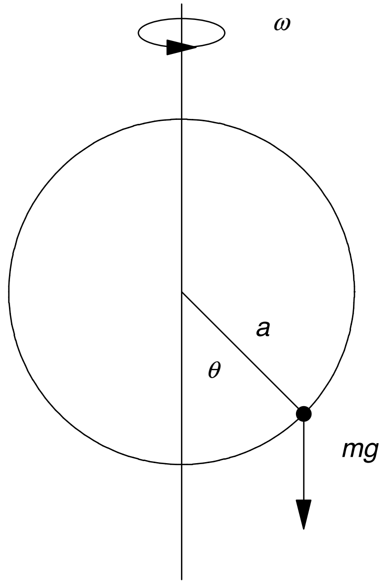



**Figure 4.8.** A rotating hoop

**4.10.6. The cantilever beam.** Solve Euler–Poisson equation (4.30) for the cantilever beam, apply the boundary conditions, and determine the deflection of the beam. How large is the deflection at the end of the beam?

**4.10.7. Beam on an elastic foundation.** Consider a cantilever beam with a uniform distributed load that rests on an elastic foundation. Assume that the elastic foundation provides a restoring force that is proportional to the displacement of the beam. Write down total potential energy for this system. Determine the Euler–Poisson equation and the boundary conditions for this problem.

**4.10.8. A rotating hoop.**

A bead of mass $m$ slides, without friction, on a circular hoop of radius $a$. The hoop lies in a vertical plane that is constrained to rotate about the hoop's vertical diameter with constant angular velocity $\omega$.

(a) Using the generalized coordinate $\theta$, determine the kinetic energy $T$, the potential energy $V$, and the Lagrangian function $L = T - V$ for the bead.

(b) Determine a first integral for this problem. Show that the first integral does not correspond to the total energy $T + V$. The total energy is evidently not constant for this problem. Discuss why this might be the case.

**4.10.9. Phase portrait for the rotating hoop.** Ignore the first integral in the (above) problem of the bead on the rotating hoop.

(a) Instead, use the Euler–Lagrange equation to write the equation of motion of the bead as a second-order differential equation.

(b) Introduce the new variable $\phi = \dot{\theta}$ and rewrite the equation of motion for the bead as two first-order differential equations (for $\dot{\theta}$ and $\dot{\phi}$).

(c) The bead is at equilibrium if $\dot{\theta} = 0$ and $\dot{\phi} = 0$. Find all possible equilibria for the bead. How do the equilibria depend on $\omega$? Discuss any bifurcations that arise as you change the angular velocity $\omega$?

(d) Draw $(\theta, \phi)$ phase portraits for your system. Be sure to draw phase portraits on each side of any bifurcation points that you have found.

**4.10.10. The pendulum revisited.** The angle $\theta$ was a convenient choice for the generalized coordinate for the simple pendulum, but it is certainly not the only possibility. Take the horizontal displacement of the mass,

$$x = l \sin \theta \,, \tag{4.161}$$

as your new generalized coordinate. Write down the kinetic energy, the potential energy, the Lagrangian, the Euler–Lagrange equation, and the energy integral for the simple pendulum in terms of this new coordinate. Your final answers may contain $m$, $g$, $l$, $x$, $\dot{x}$, and $\ddot{x}$, but not $y$ or $\theta$.

**4.10.11. Motion on a paraboloid.** Consider a particle of mass $m$ that moves, without friction, on the smooth paraboloid

$$z = x^2 + y^2 \tag{4.162}$$

while experiencing the constant force of gravity.

(a) Write down the Lagrangian for this system using the Cartesian coordinates $x$ and $y$ as your generalized coordinates. Determine the equations of motion and simplify, taking advantage of obvious first integrals.

(b) Write down the Lagrangian for this system using the cylindrical coordinates $r$ and $\theta$ as your generalized coordinates. Determine the equations of motion and simplify, taking advantage of obvious first integrals.

**4.10.12. Differentiating the Hamiltonian.** Starting with definition (4.91), show that

$$\frac{dH}{dt} = -\frac{\partial L}{\partial t} \tag{4.163}$$

along extremals.

**4.10.13. The spherical pendulum revisited.** Formulate the canonical momenta and the Hamiltonian and determine the canonical equations of motion for the spherical pendulum.

**4.10.14. Meridional rays.** Determine the shape of the meridional rays for a cylindrical optical fiber that has the refractive index

$$n^2(r) = n_0^2\,(\gamma^2 - r^2)\,. \tag{4.164}$$

The coordinate $r$ is the distance from the axis of the fiber.

**4.10.15. Triple integrals.** The Euler–Lagrange equation for double integrals generalizes easily to triple integrals. Extend the argument in the lecture notes and show that the Euler–Lagrange equation for a functional of the form

$$J[u] = \iiint_V f(x, y, z, u, u_x, u_y, u_z)\; dx\,dy\,dz \tag{4.165}$$

reduces to

$$\frac{\partial f}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial u_x}\right) - \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial u_y}\right) - \frac{\partial}{\partial z}\left(\frac{\partial f}{\partial u_z}\right) = 0\,. \tag{4.166}$$

What vector-analysis identity did you need to use to prove your result?

**4.10.16. Vibration of a drumhead.** Derive the equation for the vibration of a (two-dimensional) drumhead in both (a) Cartesian coordinates and (b) polar coordinates. Assume that the Lagrangian density of the drumhead can be written

$$\mathcal{L} = \frac{1}{2}\sigma u_t^2 - \frac{1}{2}\tau(\nabla u)^2\,, \tag{4.167}$$

where $(\nabla u)^2 \equiv (\nabla u)\cdot(\nabla u)$, $\sigma$ is the areal mass density (with units of mass per area), and $\tau$ is the surface tension or surface energy density (with dimensions of force per length or energy per area).

**4.10.17. Free oscillations of a diving board.** A uniform diving board of mass density $\rho$ and cross-sectional area $A$ has kinetic energy

$$T = \frac{\rho A}{2}\int_0^l u_t^2\, dx \tag{4.168}$$

and potential energy

$$V = \frac{EI}{2}\int_0^l u_{xx}^2\, dx\,, \tag{4.169}$$

where $E$ is Young's modulus and $I$ is the moment of inertia. Find the partial differential equation for the free oscillations of the diving board. The boundary conditions are $u = u_x = 0$ at the left end and $u_{xx} = u_{xxx} = 0$ at the right end. Find the frequency of the lowest mode of oscillation.

Chapter 5

# Constraints

## 5.1. Types of constraints

So far, we have avoided problems with side conditions. Variations may, in fact, be constrained. There are several classes of constraints that appear in the calculus of variations.

**5.1.1. Isoperimetric constraints.** In addition to the usual functional,

$$J[y] \;=\; \int_a^b f(x,y,y')\;dx\,, \tag{5.1}$$

and boundary conditions,

$$y(a) \;=\; y_a,\;\; y(b) \;=\; y_b\,, \tag{5.2}$$

we may also have one or more integral conditions of the form

$$K[y] \;=\; \int_a^b g(x,y,y')\;dx \;=\; l\,. \tag{5.3}$$

These integral conditions are called *isoperimetric constraints.*

**Example 5.1** (Queen Dido's problem)**.**

Find (see Figure 5.1), among all curves of length $l$ in the upper half-plane passing through $(-a,0)$ and $(a,0)$, the one that, together

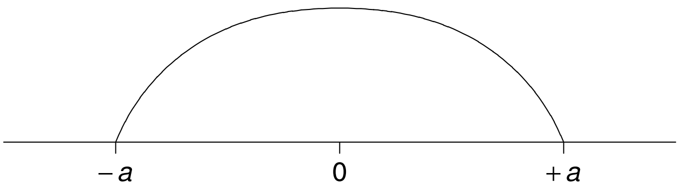



**Figure 5.1.** Queen Dido's domain?

with interval $[-a, a]$, encloses the largest area. To solve this problem, we must maximize

$$J[y] \;=\; \int_{-a}^{+a} y\; dx \tag{5.4}$$

subject to the boundary conditions

$$y(-a) \;=\; 0\,, \quad y(+a) \;=\; 0 \tag{5.5}$$

and the isoperimetric constraint

$$K[y] \;=\; \int_{-a}^{+a} \sqrt{1+y'^2}\; dx \;=\; l\,. \tag{5.6}$$

**Example 5.2** (Catenary)**.**

A heavy, uniform, and flexible chain of length $l$ hangs in equilibrium, under gravity, from two fixed points $A$ and $B$. Find the equation of the curve assumed by the chain.

In this problem, we wish to minimize the potential energy,

$$J[y] \;=\; \rho g \int_{a}^{b} y\,\sqrt{1+y'^2}\; dx\,, \tag{5.7}$$

subject to the boundary conditions

$$y(a) \;=\; y_a\,, \quad y(b) \;=\; y_b \tag{5.8}$$

and the integral condition

$$K[y] \;=\; \int_a^b \sqrt{1+y'^2}\; dx \;=\; l\,. \tag{5.9}$$

**5.1.2. Holonomic constraints.** In a variational problem with several dependent variables, a geometric restriction of the form

$$g(x, y_1, \ldots, y_n) \;=\; 0 \tag{5.10}$$

is known as a finite, positional, or *holonomic* constraint. A variational problem may have several holonomic constraints. Thus, we may wish to minimize or maximize a functional with $n$ dependent variables,

$$J[y_1, \ldots, y_n] \;=\; \int_a^b f(x,\, y_1, \ldots, y_1', \ldots, y_n')\; dx\,, \tag{5.11}$$

$2n$ boundary conditions,

$$y_i(a) \;=\; y_{ia}\,,\quad y_i(b) \;=\; y_{ib}\,,\quad i \;=\; 1, \ldots, n\,, \tag{5.12}$$

and $m$ positional constraints,

$$g_j(x, y_1, \ldots, y_n) \;=\; 0\,,\quad j \;=\; 1, \ldots, m\,. \tag{5.13}$$

Holonomic constraints are especially common in mechanics, where time is commonly the independent variable and the generalized coordinates are often the dependent variables.

**Example 5.3** (Atwood's machine)**.**

Consider two masses, connected by a string of length $l$, hung over a pulley of height $h$ (see Figure 5.2). The potential energy for this problem is

$$V \;=\; m_1 g(h - y_1) + m_2 g(h - y_2)\,. \tag{5.14}$$

The Lagrangian is thus

$$L \;=\; T - V \tag{5.15}$$

or

$$L \;=\; \frac{1}{2}(m_1\dot{y}_1^2 + m_2\dot{y}_2^2) - g[m_1(h - y_1) + m_2(h - y_2)]\,. \tag{5.16}$$

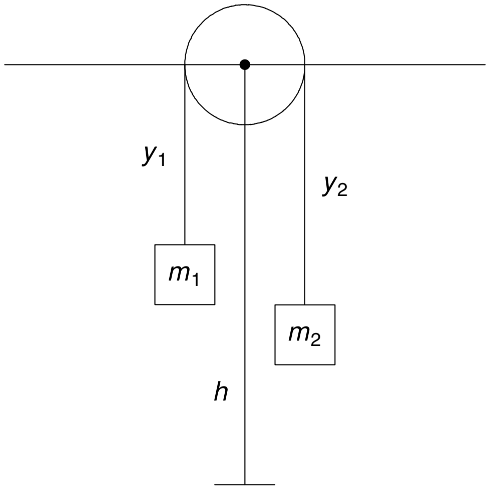



**Figure 5.2.** Atwood's machine

The functional is

$$J[y_1, y_2] \;=\; \int_{t_a}^{t_b} L(y_1, y_2, \dot{y}_1, \dot{y}_2)\; dt\,, \tag{5.17}$$

but it is subject to the holonomic constraint

$$g(y_1, y_2) \;=\; y_1 + y_2 - l \;=\; 0\,. \tag{5.18}$$

Although our functional contains both $y_1(t)$ and $y_2(t)$, our holonomic constraint implies that these two variables are not independent: our system has only one degree of freedom.

**5.1.3. Nonholonomic constraints.** We may also be forced to consider a functional with $n$ dependent variables,

$$J[y_1, \ldots, y_n] \;=\; \int_a^b f(x, y_1, \ldots, y_n, y_1', \ldots, y_n')\; dx\,, \tag{5.19}$$

$2n$ boundary conditions,

$$y_i(a) \;=\; y_{ia}\,, \quad y_i(b) \;=\; y_{ib}\,, \quad i \;=\; 1, \ldots, n\,, \tag{5.20}$$

and $m$ differential-equation constraints of the form

$$g_j(x, y_1, \ldots, y_n, y'_1, \ldots, y'_n) \;=\; 0\,, \quad j \;=\; 1, \ldots, m\,. \tag{5.21}$$

We are specifically interested in those constraints that cannot be reduced to holonomic constraints by integration. That is, we are interested in *nonholonomic* constraints.

In the simplest case, a differential-equation constraint,

$$g(x, \mathbf{y}, \mathbf{y}') \;=\; 0\,, \tag{5.22}$$

is linear in its derivatives,

$$g(x, \mathbf{y}, \mathbf{y}') \;=\; a(x, \mathbf{y}) + \sum_{i=1}^{n} b_i(x, \mathbf{y})\, y'_i \;=\; 0\,. \tag{5.23}$$

This constraint is *exact* if there exists a function $G(x, \mathbf{y})$ such that

$$\frac{dG}{dx} \;=\; \frac{\partial G}{\partial x} + \sum_{i=1}^{n} \frac{\partial G}{\partial y_i}\, y'_i \;=\; g(x, \mathbf{y}, \mathbf{y}')\,. \tag{5.24}$$

For this to be true, we clearly require

$$\frac{\partial G}{\partial x} = a(x, \mathbf{y})\,; \quad \frac{\partial G}{\partial y_i} = b_i(x, \mathbf{y})\,, \quad i \;=\; 1, \ldots, n\,. \tag{5.25}$$

The necessary and sufficient conditions for equation (5.23) to be exact are that

$$\frac{\partial a}{\partial y_i} \;=\; \frac{\partial b_i}{\partial x}\,, \quad \frac{\partial b_i}{\partial y_j} \;=\; \frac{\partial b_j}{\partial y_i}\,, \quad i, j \;=\; 1, \ldots, n\,, \tag{5.26}$$

which is equivalent to the order of differentiation not mattering in the mixed partials of $G(x, \mathbf{y})$. In general, constraint (5.23) is integrable (and hence holonomic) if it is either exact or becomes exact after multiplication by an integrating factor $\mu(x, \mathbf{y})$.

**Example 5.4** (A rolling penny)**.**

Consider a vertical penny that rolls without slipping (see Figure 5.3). Let $(x, y)$ be the coordinates of the point of contact with the surface, let $\phi$ be the angle of rotation about the axis of the penny, and let $\theta$ be the angle between the axis of the penny and the $x$-axis.

If the penny has radius $r$ and is rolling with speed

$$v \;=\; r\,\frac{d\phi}{dt}\,, \tag{5.27}$$

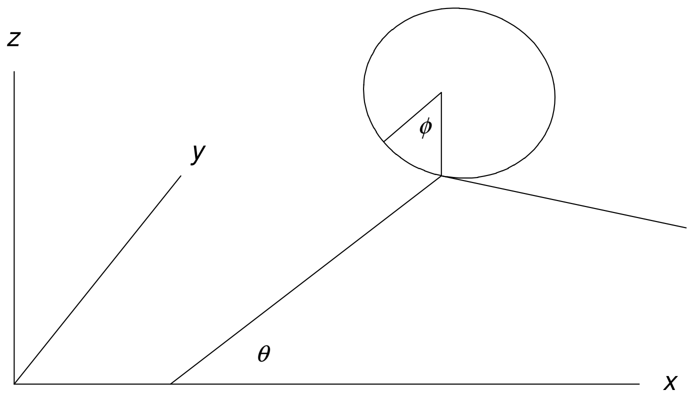



**Figure 5.3.** A rolling penny

the rate of change of the $x$ and $y$ coordinates are

$$\frac{dx}{dt} = v \sin\theta\,, \quad \frac{dy}{dt} = -v\cos\theta\,. \tag{5.28}$$

It now follows that

$$\frac{dx}{dt} - r\sin\theta\,\frac{d\phi}{dt} = 0\,, \quad \frac{dy}{dt} + r\cos\theta\,\frac{d\phi}{dt} = 0\,. \tag{5.29}$$

These are nonholonomic constraints for the four coordinates $x$, $y$, $\theta$, and $\phi$. There is no way to integrate these equations for, say, $\theta$ and $\phi$ in terms of $x$ and $y$ short of determining the actual motion of the penny. Indeed, there are many possible orientations of the coin, $\theta$ and $\phi$, for the same point of contact, $(x, y)$.

**5.1.4. One-sided or inequality constraints.** We can also consider functionals of the form

$$J[y] = \int_a^b f(x, y, y')\,dx \tag{5.30}$$

with boundary conditions

$$y(a) = y_a\,, \quad y(b) = y_b \tag{5.31}$$

and a positional inequality constraint of the form

$$g(x, y) \geq 0\,. \tag{5.32}$$

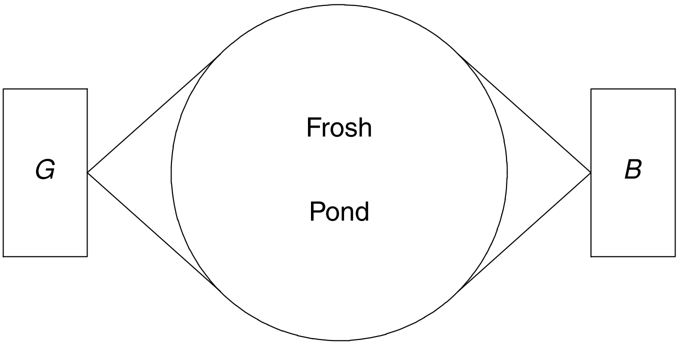



**Figure 5.4.** Mt. Rainier vista

**Example 5.5** (Frosh Pond)**.**

Consider the problem of minimizing the distance between Guggenheim and Bagley Halls on the University of Washington campus without going through Frosh Pond (see Figure 5.4). We wish to minimize the functional

$$J[y] \;=\; \int_a^b \sqrt{1+y'^2}\; dx \tag{5.33}$$

subject to the boundary conditions

$$y(a) \;=\; y_a\,, \quad y(b) \;=\; y_b \tag{5.34}$$

and the pointwise inequality constraint

$$x^2+y^2-r^2 \;\geq\; 0\,, \tag{5.35}$$

where $r$ is the radius of the pond.

It would be fun to consider all of these constraints. We do, however, have a limited amount of space and time and so we will focus on isoperimetric and holonomic constraints. Before doing so, let us first recall Lagrange multipliers.

## 5.2. Lagrange multipliers

In order to handle constraints, we will need to use Lagrange multipliers, which were developed by Lagrange to deal with problems in mechanics and the calculus of variations (Fraser, 1992). Let us recall the key geometric idea behind Lagrange multipliers by means of a simple finite-dimensional example.

**Example 5.6.**

Let us find the dimensions, $x$ and $y$, of the rectangle having the smallest perimeter among all rectangles having fixed area $A$. We clearly wish to minimize

$$f(x,y) \;=\; 2\,(x+y) \tag{5.36}$$

subject to the constraint

$$g(x,y) \;=\; xy - A \;=\; 0\,. \tag{5.37}$$

Figure 5.5 shows a graphical solution to this problem.

We want the lowest level curve of $f$ that is consistent with our constraint. This occurs at a point of tangency between the level curves of $f$ and the constraint curve $g$. At this point, the normals to both level curves are parallel,

$$\nabla f \;=\; \lambda\,\nabla g \tag{5.38}$$

so that

$$\nabla\,(f - \lambda g) \;=\; 0\,. \tag{5.39}$$

As you can imagine, $\nabla g = 0$ is a disaster for this approach since we can then no longer define the direction of the normal for the constraint curve.

Here is another, more formal, way to view the situation. Our constraint, $g(x,y)=0$, implicitly defines a curve. Let us parameterize our position on this curve with a single independent variable, say $t$. That is, let us describe our position on this curve with the smooth vector-valued function

$$\mathbf{r}(t) \;=\; [x(t), y(t)]\,. \tag{5.40}$$

We will assume that the curve is regular so that $\dot{\mathbf{r}} \neq 0$.

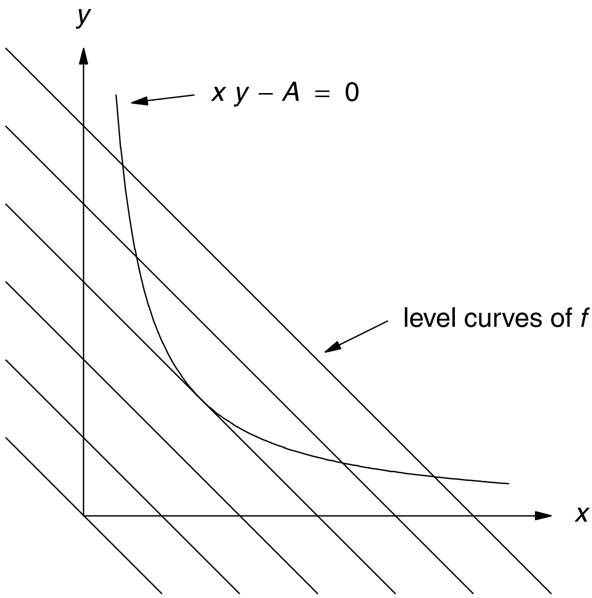



**Figure 5.5.** Geometry of Lagrange multipliers

A necessary condition for $f(x, y)$ to have local extremum along the constraint curve is

$$\frac{d}{dt} f(x(t), y(t)) \;=\; \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \;=\; 0\,. \qquad (5.41)$$

Since

$$g(x(t), y(t)) \;=\; 0\,, \qquad (5.42)$$

it also follows that

$$\frac{d}{dt} g(x(t), y(t)) \;=\; \frac{\partial g}{\partial x}\frac{dx}{dt} + \frac{\partial g}{\partial y}\frac{dy}{dt} \;=\; 0\,. \qquad (5.43)$$

If $g_x$ and $g_y$ are not both zero, i.e., if $\nabla g \neq 0$, we can thus solve for either $\dot{x}$ or $\dot{y}$ from the last displayed equation. (This is just a variant of the implicit function theorem.) Suppose, for example, that $g_y \neq 0$. Then

$$\frac{dy}{dt} \;=\; -\frac{(\partial g/\partial x)}{(\partial g/\partial y)}\frac{dx}{dt} \qquad (5.44)$$

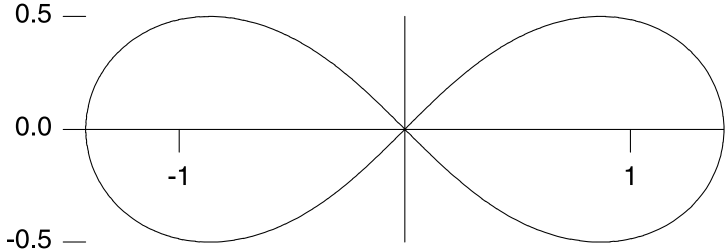



**Figure 5.6.** Lemniscate

and our necessary condition, equation (5.41), reduces to

$$\frac{dx/dt}{\partial g/\partial y}\left(\frac{\partial f}{\partial x}\frac{\partial g}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial g}{\partial x}\right) = 0\,. \tag{5.45}$$

We have assumed that our curve is regular and that $\dot{x}$ and $\dot{y}$ do not simultaneously vanish. It now follows that

$$\frac{\partial f}{\partial x}\frac{\partial g}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial g}{\partial x} = 0\,. \tag{5.46}$$

We may rewrite this last condition as

$$\nabla f \times \nabla g = \begin{vmatrix} i & j & k \\ f_x & f_y & 0 \\ g_x & g_y & 0 \end{vmatrix} = \mathbf{0}\,. \tag{5.47}$$

$\nabla f$ and $\nabla g$ are thus parallel. It follows that

$$\nabla f = \lambda \nabla g \tag{5.48}$$

and that

$$\nabla(f - \lambda g) = 0 \tag{5.49}$$

for some constant $\lambda$.

A problem with this approach occurs in those instances when $g(x, y)$ has a singular point. For example, the gradient of the lemniscate

$$g(x,y) = (x^2 + y^2)^2 - 2\,(x^2 - y^2) = 0 \tag{5.50}$$

(see Figure 5.6) is not normal to the curve at the origin. We would also find it hard to write either $y(x)$ or $x(y)$ in the immediate neighborhood of $(x, y) = (0, 0)$. We need to be careful to exclude pathologies such as this.

We will therefore assume that $f(x, y)$ and $g(x, y)$ are differentiable in a neighborhood of the points of the regular curve $g(x, y) = 0$ and that at least one partial derivative of $g$, either $g_x$ or $g_y$, is nonzero at every point of the curve $g(x, y) = 0$. As a result, $\nabla g \neq 0$ along the curve.

With these assumptions, we now conclude:

***Lagrange multiplier rule:***

If the function $z = f(x, y)$ has a relative extremum on the curve $g(x, y) = 0$ at the point $(x_0, y_0)$, then there exists a constant $\lambda$ and a function

$$F(x, y) \equiv f(x, y) - \lambda\, g(x, y) \tag{5.51}$$

such that

$$\frac{\partial F}{\partial x}(x_0, y_0) = 0\,, \quad \frac{\partial F}{\partial y}(x_0, y_0) = 0 \tag{5.52}$$

and

$$g(x_0, y_0) = 0\,. \tag{5.53}$$

We now need to generalize this result to the calculus of variations. This can happen in a number of different ways depending on the nature of the constraint.

## 5.3. Isoperimetric constraints

Let us start with the simplest problem, one with an integral constraint. We are interested in finding the extremum of the functional

$$J[y] = \int_a^b f(x, y, y')\; dx \tag{5.54}$$

subject to boundary conditions

$$y(a) = y_a\,, \quad y(b) = y_b \tag{5.55}$$

and the added restriction that

$$K[y] \;=\; \int_a^b g(x,y,y')\;dx \;=\; l\,. \tag{5.56}$$

We will, as usual, assume that there is a solution and we will embed the assumed local extremum, $\hat{y}(x)$, in a family of varied curves.

An arbitrary one-parameter family of variations will not, by itself, work since the resulting curves may not satisfy the integral constraint. We will instead give ourselves some extra freedom and introduce the *two-parameter* family

$$y(x) \;=\; \hat{y}(x) + \epsilon_1\,\eta_1(x) + \epsilon_2\,\eta_2(x)\,. \tag{5.57}$$

The weak variations $\eta_1(x)$ and $\eta_2(x)$ are independent and arbitrary save for the fact that

$$\eta_1(a) \;=\; \eta_2(a) \;=\; 0\,, \quad \eta_1(b) \;=\; \eta_2(b) \;=\; 0\,. \tag{5.58}$$

Constraint equation (5.56) is satisfied at $\epsilon_1 = \epsilon_2 = 0$, by assumption. We can think of $\epsilon_1$ as small, but arbitrary, and $\epsilon_2$ as a "correction term," guaranteed by the implicit function theorem, that ensures that the integral constraint is satisfied. Alternatively, we can think of $\epsilon_2$ as small and arbitrary and $\epsilon_1$ as the correction term. Either way, because $\epsilon_1$ and $\epsilon_2$ are related by our constraint equation, we ultimately have only one free and arbitrary parameter.

The two parameters $\epsilon_1$ and $\epsilon_2$ allow us to work with functions instead of functionals. Evaluating our two functionals along the specified $y(x)$ gives us, in effect, the functions

$$J(\epsilon_1,\epsilon_2) \;=\; \int_a^b f(x,y,y')\;dx\,, \tag{5.59}$$

$$K(\epsilon_1,\epsilon_2) \;=\; \int_a^b g(x,y,y')\;dx \;=\; l\,. \tag{5.60}$$

This, however, is just our old finite-dimensional calculus problem; we may now apply Lagrange multipliers.

Let

$$I \;\equiv\; J - \lambda\,K\,, \quad F \;\equiv\; f - \lambda\,g\,. \tag{5.61}$$

Since $\epsilon_1 = \epsilon_2 = 0$ corresponds to our supposed extremum, there should be a ***constant*** $\lambda$ such that

$$\frac{\partial I}{\partial \epsilon_1} = \frac{\partial I}{\partial \epsilon_2} = 0 \tag{5.62}$$

at $(\epsilon_1, \epsilon_2) = (0, 0)$. We expect this to be true as long as the $\partial K / \partial \epsilon_i$ do not both vanish.

For the functional $I$,

$$\begin{aligned} \frac{\partial I}{\partial \epsilon_i} &= \int_a^b \left( \frac{\partial F}{\partial y} \frac{\partial y}{\partial \epsilon_i} + \frac{\partial F}{\partial y'} \frac{\partial y'}{\partial \epsilon_i} \right) dx \\ &= \int_a^b \left( \frac{\partial F}{\partial y} \eta_i + \frac{\partial F}{\partial y'} \eta_i' \right) dx \,. \end{aligned} \tag{5.63}$$

After integrating by parts, we now have

$$\frac{\partial I}{\partial \epsilon_i} = \int_a^b \left[ \frac{\partial F}{\partial y} - \frac{d}{dx} \left( \frac{\partial F}{\partial y'} \right) \right] \eta_i(x) \; dx \,. \tag{5.64}$$

Applying

$$\frac{\partial I}{\partial \epsilon_1} = \frac{\partial I}{\partial \epsilon_2} = 0 \tag{5.65}$$

at

$$(\epsilon_1, \epsilon_2) = (0, 0) \,, \tag{5.66}$$

it now follows that

$$\left\{ \frac{\partial (f - \lambda g)}{\partial y} - \frac{d}{dx} \left[ \frac{\partial (f - \lambda g)}{\partial y'} \right] \right\}_{\hat{y}, \hat{y}'} = 0 \,. \tag{5.67}$$

In effect, all we have to do is to consider $f - \lambda g$, with $\lambda$ constant, and proceed as before. More precisely, we will follow Euler's (isoperimetric) rule (Pars, 1962; Clegg, 1968):

(1) First find the extremals for the integral

$$\int_a^b [f(x, y, y') - \lambda \, g(x, y, y')] \; dx \,, \tag{5.68}$$

for constant $\lambda$, in the form

$$y = y(x, \lambda, c_1, c_2)\,, \tag{5.69}$$

where $c_1$ and $c_2$ are constants.

(2) Then choose $\lambda$, $c_1$, and $c_2$ so that

(a) the extremal satisfies the boundary conditions

$$y(a) = y_a\,, \quad y(b) = y_b\,, \tag{5.70}$$

(b) the extremal gives $K$ the value $l$.

There are thus three conditions for the three constants $\lambda$, $c_1$, and $c_2$. In general, we expect the above method to work as long $y = \hat{y}(x)$ is not an extremal of $K[y]$.

## 5.4. Case study: Queen Dido's problem

Let us reconsider the problem of maximizing the area

$$J[y] = \int_{-a}^{+a} y\,dx \tag{5.71}$$

subject to the boundary conditions

$$y(-a) = 0\,, \quad y(+a) = 0 \tag{5.72}$$

and the integral constraint

$$K[y] = \int_{-a}^{+a} \sqrt{1 + y'^2}\,dx = l\,. \tag{5.73}$$

There is only one constraint in this problem and so we expect one Lagrange multiplier. So, let us introduce

$$J - \lambda K = \int_{-a}^{+a} y - \lambda\sqrt{1 + y'^2}\,dx\,. \tag{5.74}$$

The variable $x$ is not explicitly present and we could write down a first integral, but this problem proceeds equally smoothly — try it both ways — if we write down the full Euler–Lagrange equation

$$1 + \lambda \frac{d}{dx}\left(\frac{y'}{\sqrt{1+y'^2}}\right) \;=\; 0\,, \tag{5.75}$$

which immediately integrates to

$$\frac{\lambda y'}{\sqrt{1+y'^2}} \;=\; -(x-c_1)\,. \tag{5.76}$$

Solving for $y'$ and separating variables produces

$$dy \;=\; \pm\frac{(x-c_1)}{\sqrt{\lambda^2-(x-c_1)^2}}\,dx\,, \tag{5.77}$$

which integrates out to

$$y \;=\; \pm\sqrt{\lambda^2-(x-c_1)^2}\;+\;c_2\,. \tag{5.78}$$

This last equation simplifies to the equation of a circle with center $(c_1, c_2)$ and radius $\lambda$,

$$(x-c_1)^2+(y-c_2)^2 \;=\; \lambda^2\,. \tag{5.79}$$

Given the symmetry of the problem, it is easy to show that $c_1 = 0$. To determine $c_2$ and $\lambda$, we may look at the geometry of the problem. From Figure 5.7,

$$\lambda \sin\theta \;=\; a\,, \quad \lambda\,\theta \;=\; \frac{l}{2} \tag{5.80}$$

so that

$$\sin\theta \;=\; \frac{2a}{l}\,\theta\,. \tag{5.81}$$

This transcendental equation can be solved graphically (see Figure 5.8). With $\theta$ in hand,

$$\lambda \;=\; a\csc\theta\,, \quad c_2 \;=\; a\cot\theta\,. \tag{5.82}$$

We will return to this problem when we talk about parametric problems.

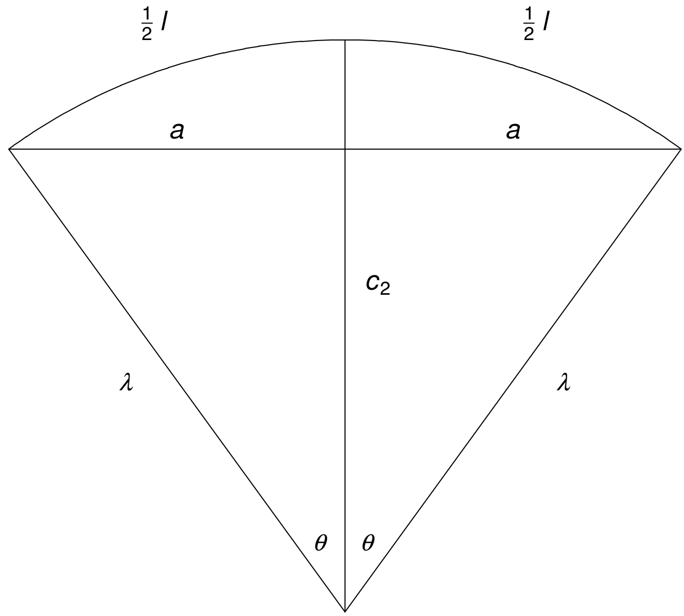



**Figure 5.7.** Geometry of Queen Dido's problem

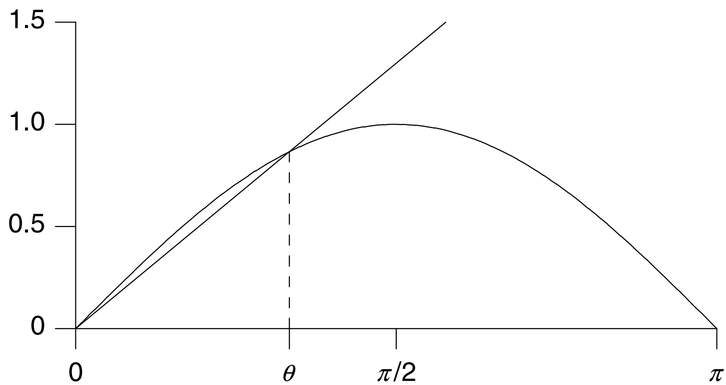



**Figure 5.8.** Graphical solution of $u$

## 5.5. Case study: Euler's elastica

Isoperimetric constraints also arise in the study of functionals with higher-order derivatives, such as those that arise in solid mechanics. In his 1744 book, Euler tackled the problem of minimizing the elastic strain energy of an elastic rod of fixed length with fixed boundary conditions.

In our discussion of the cantilever beam in Chapter 4, we assumed that beam deflections were small, so that we could neglect second-degree terms in $y'(x)$. Euler chose to keep these second-degree terms so that he could study large deflections. He instead dropped the effects of gravity. This, in turn, allowed him to ignore several parameters. He thus considered the problem of minimizing the functional

$$J[y] \;=\; \int_a^b f(x,y,y',y'')\;dx \;=\; \int_a^b \frac{(y'')^2}{(1+y'^2)^{5/2}}\;dx \tag{5.83}$$

subject to the constraint

$$K[y] \;=\; \int_a^b g(x,y,y')\;dx \;=\; \int_a^b \sqrt{1+y'^2}\;dx \;=\; l \tag{5.84}$$

and the boundary conditions

$$y(a) \;=\; y_a\,,\quad y'(a) \;=\; y'_a\,,\quad y(b) \;=\; y_b\,,\quad y'(b) \;=\; y'_b\,. \tag{5.85}$$

We may now proceed in the usual manner. Let

$$I \;\equiv\; J - \lambda\,K \tag{5.86}$$

and

$$\begin{aligned} F \;&\equiv\; f(x,y,y',y'') - \lambda\,g(x,y,y') \\ &=\; \frac{(y'')^2}{(1+y'^2)^{5/2}} - \lambda\sqrt{1+y'^2}\,. \end{aligned} \tag{5.87}$$

Since our integrand contains a second derivative, we expect our Euler–Poisson equation,

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial y''}\right) \;=\; 0\,, \tag{5.88}$$

to produce a fourth-order differential equation. Euler made quick work of this fourth-order differential equation by integrating it three times. Let us follow his path in doing so.

Since our integrand $F(x, y, y', y'')$ does not contain $y(x)$, we can now write our Euler–Poisson equation as

$$-\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial f}{\partial y''}\right) + \lambda\,\frac{d}{dx}\left(\frac{\partial g}{\partial y'}\right) \;=\; 0\,, \tag{5.89}$$

which, after integration, yields

$$P - \frac{dQ}{dx} - \lambda\,\frac{y'}{\sqrt{1+y'^2}} \;=\; \beta\,, \tag{5.90}$$

where

$$P \;\equiv\; \frac{\partial f}{\partial y'} \;=\; -\frac{5y'(y'')^2}{(1+y'^2)^{7/2}} \tag{5.91}$$

and

$$Q \;\equiv\; \frac{\partial f}{\partial y''} \;=\; \frac{2y''}{(1+y'^2)^{5/2}}\,. \tag{5.92}$$

One integration down.

Now, this is probably as far as most mathematicians would have gotten with this problem. Fortunately, Euler had a few more tricks up his sleeves. Multiplying equation (5.90) by

$$dy' \;=\; y''\,dx \tag{5.93}$$

produces

$$P\,dy' - y''\,dQ - \lambda\,\frac{y'\,dy'}{\sqrt{1+y'^2}} \;=\; \beta\,dy'\,. \tag{5.94}$$

Since

$$df \;=\; P\,dy' + Q\,dy''\,, \tag{5.95}$$

we may add

$$df - P\,dy' - Q\,dy'' \;=\; 0 \tag{5.96}$$

to equation (5.94) to obtain

$$df - (Q\,dy'' + y''\,dQ) \;=\; \lambda\frac{y'\,dy'}{\sqrt{1+y'^2}} + \beta\,dy'\,. \tag{5.97}$$

Integrating this last equation produces

$$f - y''\,Q \;=\; \lambda\sqrt{1+y'^2} + \beta\,y' + \gamma\,, \tag{5.98}$$

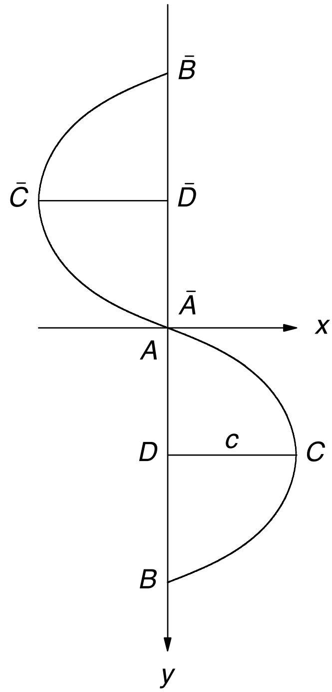



**Figure 5.9.** An elastica of class 2

where $\gamma$ is a new constant of integration. Evaluating the left-hand side of this last equation produces

$$-\frac{(y'')^2}{(1+y'^2)^{5/2}} = \lambda\sqrt{1+y'^2} + \beta\, y' + \gamma \tag{5.99}$$

while solving for $y''$ (while flipping the signs of $\lambda$, $\beta$, and $\gamma$) now yields

$$y'' = \frac{dy'}{dx} = (1+y'^2)^{5/4}\left(\lambda\sqrt{1+y'^2} + \beta\, y' + \gamma\right)^{1/2}. \tag{5.100}$$

Two integrations down.

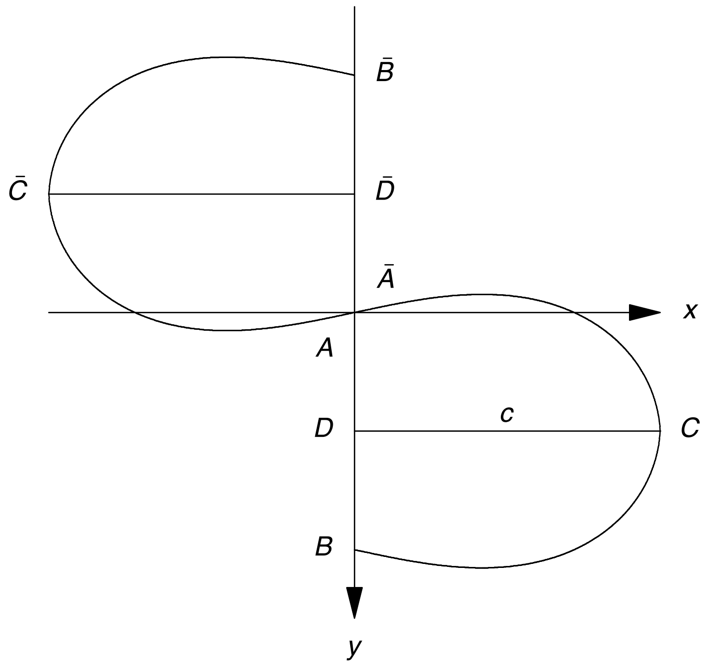



**Figure 5.10.** An elastica of class 4

Euler now observed that

$$\frac{d}{dy'}\left[\frac{2\,(\lambda\sqrt{1+y'^2}+\beta\,y'+\gamma)^{1/2}}{(1+y'^2)^{1/4}}\right] \tag{5.101}$$
$$= \frac{\beta-\gamma\,y'}{(1+y'^2)^{5/4}(\lambda\sqrt{1+y'^2}+\beta\,y'+\gamma)^{1/2}}$$

so that, using equation (5.100),

$$\frac{d}{dy'}\left[\frac{2\,(\lambda\sqrt{1+y'^2}+\beta\,y'+\gamma)^{1/2}}{(1+y'^2)^{1/4}}\right] = (\beta-\gamma y')\,\frac{dx}{dy'}\,. \tag{5.102}$$

Both sides can now be integrated to produce

$$\frac{2\,(\lambda\sqrt{1+y'^2}+\beta\,y'+\gamma)^{1/2}}{(1+y'^2)^{1/4}} = \beta x-\gamma y+\delta\,, \tag{5.103}$$

where $\delta$ is a new constant of integration. Three integrations down.

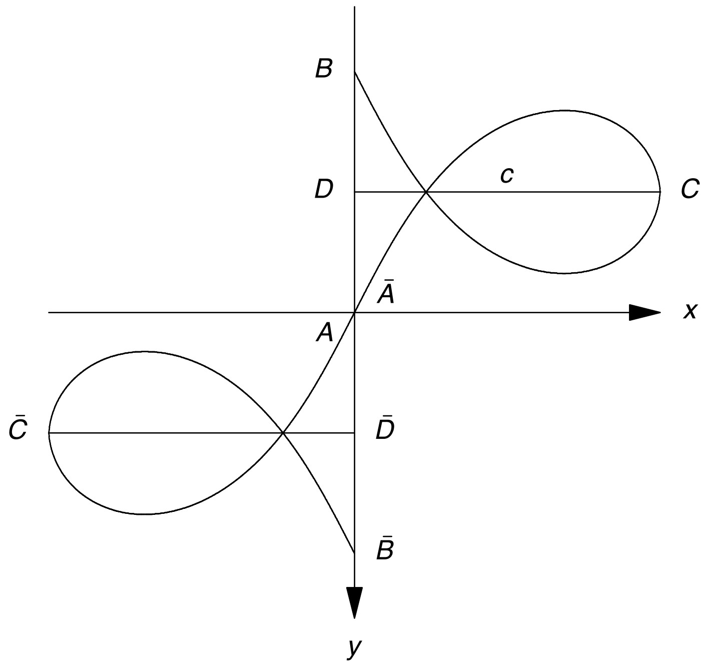



**Figure 5.11.** An elastica of class 6

Rotating and translating the coordinate system allows us to choose a coordinate system in which $\gamma = 0$ and $\delta = 0$. With this simplification, we can now solve for $y'(x)$ to obtain

$$y'(x) \;=\; \frac{\beta^2 x^2 - 4\lambda}{\sqrt{16\beta^2 - (\beta^2 x^2 - 4\lambda)^2}}\,. \tag{5.104}$$

The substitutions

$$\lambda = \frac{4m}{a^2}\,, \quad \beta = \frac{4n}{a^2} \tag{5.105}$$

give us the slightly tidier

$$y'(x) \;=\; \frac{n^2 x^2 - m a^2}{\sqrt{n^2 a^4 - (n^2 x^2 - m a^2)^2}}\,. \tag{5.106}$$

Euler then moved the coordinates back into a general position so that

$$y'(x) = \frac{\alpha + \beta x + \gamma x^2}{\sqrt{a^4 - (\alpha + \beta x + \gamma x^2)^2}}, \tag{5.107}$$

where $\alpha$, $\beta$, and $\gamma$ are now new constants. With this $y'(x)$ in hand, we can also write a differential equation for the arc length $s(x)$ of the curve,

$$s'(x) = \sqrt{1 + y'^2} = \frac{a^2}{\sqrt{a^4 - (\alpha + \beta x + \gamma x^2)^2}}. \tag{5.108}$$

The above equations yield elliptic integrals, but elliptic functions were unknown in 1744. Even so, Euler proceeded to classify the many curves that emerge from the above equations. To do so, Euler took a version of equation (5.107),

$$y'(x) = \frac{\alpha + x^2}{\sqrt{a^4 - (\alpha + x^2)^2}}, \tag{5.109}$$

where the origin (see Figure 5.9) is at $A$ and the $y$-axis is downwards. He also set

$$(a^2 - \alpha) = c^2 \tag{5.110}$$

so that

$$y'(x) = \frac{a^2 - c^2 + x^2}{\sqrt{(c^2 - x^2)(2a^2 - c^2 + x^2)}}, \tag{5.111}$$

$$s'(x) = \frac{a^2}{\sqrt{(c^2 - x^2)(2a^2 - c^2 + x^2)}}. \tag{5.112}$$

With these derivatives, it is clear that the curve $y(x)$ lies between $x = \pm c$ and that the slope of $y(x)$ at $x = \pm c$ is infinite. One can now sketch the portions $AC$ and $\overline{AC}$ of the elastica using the formula for $y'(x)$. Euler, moreover, showed that $CB$ is a reflection of $CA$ about the line $CD$ and that $\overline{CB}$ is a reflection of $\overline{CA}$ about $\overline{CD}$. This allowed Euler to fill in the rest of the periodic elastica.

The form of the elastica depends on the precise values of $c$ and $a$. Euler identified nine classes of solution curves. Figures 5.9, 5.10, and 5.11 are examples of classes 2, 4, and 6. Class 2 occurs for $0 < c < a$. For this set of $c$ values, the elastica appears sinusoidal. Class 4 occurs for $a < c < a\sqrt{1.651868}$. For $c = a\sqrt{1.651868}$, the points $A$ and $B$ coincide and the elastica becomes a figure eight. This figure eight (lemnoid or lemniscoid) is Euler's class 5. Finally, class 6 occurs for $a\sqrt{1.651868} < c < a\sqrt{2}$. In this case, point $B$ has crossed past point $A$.

## 5.6. Holonomic constraints

Let us now consider a functional

$$J[y,z] \;=\; \int_a^b f(x,y,z,y',z')\;dx \tag{5.113}$$

with two dependent variables $y(x)$ and $z(x)$, boundary conditions

$$y(a) \;=\; y_a\,,\quad y(b) \;=\; y_b\,,\quad z(a) \;=\; z_a\,,\quad z(b) \;=\; z_b\,, \tag{5.114}$$

and a simple holonomic constraint

$$g(x,y,z) \;=\; 0\,. \tag{5.115}$$

We will assume that there are two functions, $\hat{y}(x)$ and $\hat{z}(x)$, that minimize the functional for the given conditions. If we perturb $\hat{y}(x)$,

$$y(x) \;=\; \hat{y}(x) + \epsilon\,\eta_1(x)\,, \tag{5.116}$$

with an $\eta_1(x)$ that satisfies

$$\eta_1(a) \;=\; \eta_1(b) \;=\; 0\,, \tag{5.117}$$

$\hat{z}(x)$ will also experience a perturbation,

$$z(x) \;=\; \hat{z}(x) + \eta_2(x,\epsilon)\,, \tag{5.118}$$

because of the positional constraint. In other words, the variations in $y$ and $z$ are not independent.

One can easily show that

$$\eta_2(x,0) \;=\; 0 \tag{5.119}$$

and that

$$\eta_2(a,\epsilon) \;=\; \eta_2(b,\epsilon) \;=\; 0\,. \tag{5.120}$$

In addition, because of our holonomic constraint,

$$g(x,\hat{y}+\epsilon\,\eta_1,\hat{z}+\eta_2) \;=\; 0. \tag{5.121}$$

If we differentiate this equation with respect to $\epsilon$, we also find that

$$\frac{\partial g}{\partial y}\,\eta_1 + \frac{\partial g}{\partial z}\,\frac{\partial \eta_2}{\partial \epsilon} \;=\; 0\,, \tag{5.122}$$

which will soon prove helpful. This equation is true in general and we certainly expect it to be true at $\epsilon = 0$.

By assumption, $J[y, z]$ has an extremum for $\epsilon = 0$. Let us therefore examine

$$\left[\frac{d}{d\epsilon}\int_a^b f(x, y, z, y', z')\;dx\right]_{\epsilon=0} \;=\; 0\,. \tag{5.123}$$

We now get

$$\int_a^b \left[\left(\frac{\partial f}{\partial y}\,\eta_1 + \frac{\partial f}{\partial y'}\,\eta_1'\right) + \left(\frac{\partial f}{\partial z}\,\frac{\partial \eta_2}{\partial \epsilon} + \frac{\partial f}{\partial z'}\,\frac{\partial \eta_2'}{\partial \epsilon}\right)\right]_{\epsilon=0} dx \;=\; 0\,. \tag{5.124}$$

After the usual integration by parts,

$$\int_a^b \left\{\left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right]\eta_1 + \left[\frac{\partial f}{\partial z} - \frac{d}{dx}\left(\frac{\partial f}{\partial z'}\right)\right]\cdot\frac{\partial \eta_2}{\partial \epsilon}\right\}_{\epsilon=0} dx \;=\; 0\,. \tag{5.125}$$

That is all that we can say unless we can somehow remove the troublesome $\partial\eta_2/\partial\epsilon$. Fortunately, because of equation (5.122), we can solve for $\partial\eta_2/\partial\epsilon$ (at $\epsilon = 0$). Indeed, as long as $g_z \equiv \partial g/\partial z \neq 0$ along our extremal,

$$\frac{\partial \eta_2}{\partial \epsilon} \;=\; -\frac{g_y}{g_z}\,\eta_1\,. \tag{5.126}$$

(If $g_z = 0$, but $g_y \equiv \partial g/\partial y \neq 0$, we can reverse the role of $\eta_1$ and $\eta_2$. If both partial derivatives vanish along our extremal, we are out of luck.)

It now follows that

$$\int_a^b \eta_1(x)\left[f_y - \frac{d}{dx}f_{y'} - \frac{g_y}{g_z}\left(f_z - \frac{d}{dx}f_{z'}\right)\right]_{\hat{y},\hat{y}',\hat{z},\hat{z}'} dx \;=\; 0\,. \tag{5.127}$$

The subscript on the right square bracket in this last equation signifies that the expression in square brackets is evaluated at $y = \hat{y}(x)$, $y' =$

$\hat{y}'(x)$, $z = \hat{z}(x)$, and $z' = \hat{z}'(x)$. Since $\eta_1(x)$ is arbitrary, the equations

$$f_y - \frac{d}{dx} f_{y'} - \frac{g_y}{g_z}\left(f_z - \frac{d}{dx} f_{z'}\right) \;=\; 0 \tag{5.128}$$

and

$$\frac{f_y - \frac{d}{dx} f_{y'}}{g_y} \;=\; \frac{f_z - \frac{d}{dx} f_{z'}}{g_z} \tag{5.129}$$

must be satisfied along an extremal.

The common value of the above ratios is some function of $x$. If we denote this function by $\lambda(x)$, it now follows that

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) \;=\; 0\,, \quad \frac{\partial F}{\partial z} - \frac{d}{dx}\left(\frac{\partial F}{\partial z'}\right) \;=\; 0\,, \tag{5.130}$$

where

$$F \;=\; f(x, y, z, y', z') - \lambda(x)\, g(x,\, y,\, z)\,. \tag{5.131}$$

Equations (5.130) provide two differential equations for the three unknown functions $\hat{y}(x)$, $\hat{z}(x)$, and $\lambda(x)$. The constraint equation

$$g(x, \hat{y}, \hat{z}) \;=\; 0 \tag{5.132}$$

provides the third equation. Note that our Lagrange multiplier $\lambda(x)$ is now a function of $x$. Our holonomic constraint gives us a $\lambda$ at each value of $x$.

Sometimes the meaning of the Lagrange multiplier is unimportant. In other problems, it is of critical importance. In mechanics, there is a special meaning that is usually attached to the Lagrange multiplier in the case of holonomic constraints. Note that for a Lagrangian,

$$L(t, \mathbf{q}, \dot{\mathbf{q}}) \;=\; T(t, \mathbf{q}, \dot{\mathbf{q}}) - V(t, \mathbf{q})\,, \tag{5.133}$$

with a holonomic constraint equation, the $i$th Euler–Lagrange equation looks like

$$\frac{\partial L}{\partial q_i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \lambda(t)\,\frac{\partial g}{\partial q_i} \;=\; 0\,. \tag{5.134}$$

The last term on the left-hand side looks much like the force, $-\partial V/\partial q_i$, due to a potential. As a result, the last term on the left-hand side of equation (5.134) is often thought of as the constraint force in the $q_i$ direction.

Lagrange multipliers generalize naturally to problems with several isoperimetric or holonomic constraints and, in addition, to problems with nonholonomic constraints. A notorious exception occurs in mechanics. Hamilton's principle was originally derived for holonomic constraints (using d'Alembert's principle), and Hertz (1899), in a seminal work, showed that Hamilton's principle, as generally written, is not valid for nonholonomic constraints. Hölder fixed the problem, but at the cost of reformulating Hamilton's principle so that it is no longer a problem in the calculus of variations. See Capon (1952), Jeffreys (1954), Pars (1954), Rumiantsev (1982), Flannery (2005), and Lützen (2005) for further details and discussion.

It is also worth noting that an isoperimetric constraint,

$$K[y] \;=\; \int_a^b g(x,y,y')\;dx \;=\; l\,, \tag{5.135}$$

can be rewritten as a differential-equation constraint,

$$z' \;=\; g(x,y,y')\,, \tag{5.136}$$

with the new variable $z$ and the boundary conditions

$$z(a) \;=\; 0\,, \quad z(b) \;=\; l\,. \tag{5.137}$$

The corresponding $F$ is then just

$$F \;\equiv\; f - \lambda(x)\,[z' - g(x,y,y')] \tag{5.138}$$

and the corresponding Euler–Lagrange equation in $z$,

$$\frac{\partial F}{\partial z} - \frac{d}{dx}\left(\frac{\partial F}{\partial z'}\right) \;=\; 0\,, \tag{5.139}$$

implies that

$$\frac{d\lambda}{dx} \;=\; 0 \tag{5.140}$$

or that

$$\lambda \;=\; \text{constant}\,. \tag{5.141}$$

Before leaving constraints, let us work though a detailed example of a mechanical system with a holonomic constraint.

## 5.7. Case study: A sliding rod

Consider two particles, of mass $m$, connected by a rigid but massless rod of length $l$ (see Figure 5.12). We will assume that both particles move without friction. Particle $A$ moves along the $y$-axis, while particle $B$ moves along the $x$-axis. We wish to derive the differential equations of motion for the two particles. We may derive these equations in three different ways.

*Method 1:*

Let us first use the holonomic constraint

$$g(x,y) \;=\; x^2 + y^2 - l^2 \;=\; 0 \tag{5.142}$$

as an explicit constraint. The kinetic energy is

$$T \;=\; \frac{1}{2}\, m\, (\dot{x}^2 + \dot{y}^2)\,, \tag{5.143}$$

the potential energy is

$$V \;=\; mgy\,, \tag{5.144}$$

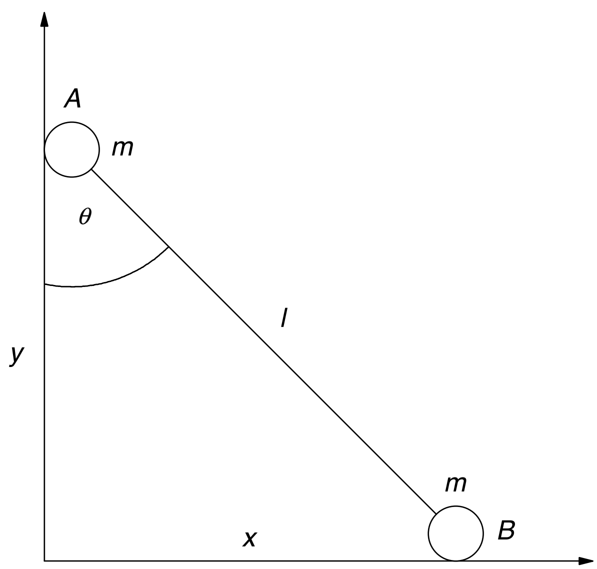



**Figure 5.12.** A sliding rod

and the Lagrangian is

$$L = T - V = \frac{1}{2} m\,(\dot{x}^2 + \dot{y}^2) - mgy\,. \tag{5.145}$$

We now form the new function

$$F = L - \lambda(t)\, g(x, y)\,, \tag{5.146}$$

which takes the form

$$F = \frac{1}{2} m\,(\dot{x}^2 + \dot{y}^2) - mgy - \lambda(t)\,(x^2 + y^2 - l^2)\,. \tag{5.147}$$

Our two Euler–Lagrange equations,

$$\frac{\partial F}{\partial x} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}}\right) = 0 \quad \text{and} \quad \frac{\partial F}{\partial y} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{y}}\right) = 0\,, \tag{5.148}$$

simplify to

$$m\ddot{x} = -2\lambda x \qquad m\ddot{y} = -2\lambda y - mg\,. \tag{5.149}$$

We also need to use our constraint equation. Differentiating our constraint equation once,

$$2x\dot{x} + 2y\dot{y} = 0\,, \tag{5.150}$$

and then twice, we get, after dividing by two,

$$x\,\ddot{x} + y\,\ddot{y} + \dot{x}^2 + \dot{y}^2 = 0\,. \tag{5.151}$$

Substituting our accelerations and solving for $\lambda$ produces

$$\lambda = \frac{m}{2l^2}[-gy + (\dot{x}^2 + \dot{y}^2]\,. \tag{5.152}$$

This expression for $\lambda$ may be substituted back into our two Euler–Lagrange equations to produce

$$\ddot{x} = \frac{x}{l^2}\,[gy - (\dot{x}^2 + \dot{y}^2)] \tag{5.153}$$

and

$$\ddot{y} = \frac{y}{l^2}\,[gy - (\dot{x}^2 + \dot{y}^2)] - g\,. \tag{5.154}$$

*Method 2:*

A second approach is to work with a single generalized coordinate that has the constraint built in. The angle $\theta$ is an obvious choice. Since

$$x = l\,\sin\theta\,, \quad y = l\,\cos\theta\,, \tag{5.155}$$

$$\dot{x} = l\,\cos\theta\,\dot{\theta}\,, \quad \dot{y} = -l\,\sin\theta\,\dot{\theta}\,,$$

the kinetic energy is

$$T = \frac{1}{2}\, m\, (\dot{x}^2 + \dot{y}^2) = \frac{1}{2}\, m\, l^2\, \dot{\theta}^2 \,, \tag{5.156}$$

the potential energy is

$$V = mgy = mgl \cos\theta \,, \tag{5.157}$$

and the Lagrangian is

$$L = \frac{1}{2}\, m\, l^2\, \dot{\theta}^2 - mgl \cos\theta \,. \tag{5.158}$$

The single Euler–Lagrange equation,

$$\frac{\partial L}{\partial \theta} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) = 0 \,, \tag{5.159}$$

now reduces to

$$ml^2 \ddot{\theta} - mgl \sin\theta = 0 \,. \tag{5.160}$$

*Method 3:*

A third approach is to use the constraint equation to eliminate one of the dependent variables. We may, for example, eliminate $x$ and its derivatives. Thus

$$x = \sqrt{l^2 - y^2} \tag{5.161}$$

and

$$\dot{x} = \frac{-y\dot{y}}{\sqrt{l^2 - y^2}} \,. \tag{5.162}$$

The kinetic energy is now

$$T = \frac{1}{2}\, m\, \dot{y}^2 \left(\frac{y^2}{l^2 - y^2} + 1\right) = \frac{1}{2}\, m\, \dot{y}^2 \left(\frac{l^2}{l^2 - y^2}\right) \tag{5.163}$$

while the potential energy is just

$$V = mgy \,. \tag{5.164}$$

A little bit of effort now leads to the differential equation

$$ml^2 \left[\frac{\ddot{y}}{l^2 - y^2} + \frac{y\dot{y}^2}{(l^2 - y^2)^2}\right] + mg = 0 \,. \tag{5.165}$$

The second method gave us the tidiest solution. However, the first method provided us additional information about the compressive force in the rod, $2\lambda l$, with $x$ and $y$ components $2\lambda x$ and $2\lambda y$, that is lost in the second method.

## 5.8. Recommended reading

Fraser (1992) reviews the early history of isoperimetric problems in the calculus of variations.

Oldfather et al. (1933) translated Euler's famous appendix on elastic curves into English. Truesdell (1960) and Heyman (1996) provide detailed descriptions of Euler's analysis of the elastica. Finally, Truesdell (1983), Fraser (1991), Heyman (1998), D'Antonio (2007), and Goss (2009) put the elastica into historical context.

Bliss (1930) is a good starting point for more on nonholonomic constraints and the calculus of variations.

See Petrov (1968) or Smith (1974) for an introduction to variational problems with inequality constraints. Problems with inequality constraints can often be turned into problems with equality constraints using slack variables (Valentine, 1937).

## 5.9. Exercises

**5.9.1. A simple constraint.** Find the curve $y(x)$ that minimizes

$$J = \frac{1}{2} \int_0^1 y'^2 \, dx \tag{5.166}$$

subject to the constraint

$$K = \int_0^1 y \, dx = \frac{1}{6} \tag{5.167}$$

and the boundary conditions

$$y(0) = 0 \quad \text{and} \quad y(1) = 0 \,. \tag{5.168}$$

**5.9.2. A hanging chain.** A heavy, uniform, and flexible chain of linear density $\rho$, shape $y(x)$, and length

$$l = \int_{-a}^{+a} \sqrt{1 + y'^2} \, dx \tag{5.169}$$

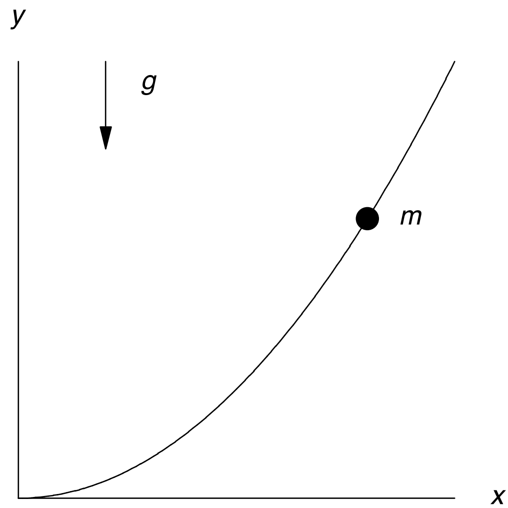



**Figure 5.13.** A parabolic wire

hangs in equilibrium, under the force of gravity, between the two points $A = (-a, h)$ and $B = (a, h)$. Determine $y(x)$. What can you say about the constants of integration?

### 5.9.3. Particle on a parabolic wire.

A particle of mass $m$ slides down a parabolic wire under the action of gravity. The equation for the wire is $y = x^2$.

(a) Determine the equations of motion for this system using Lagrange multipliers.

(b) Determine the equation of motion for this system using the constraint equation to eliminate $y$ from the problem.

### 5.9.4. Atwood's machine.

(a) Determine the equations of motion (the equations for $\ddot{y}_1$ and $\ddot{y}_2$) for Atwood's machine using Lagrange multipliers. Determine and interpret the Lagrange multiplier $\lambda$.

(b) Determine the equation for $\ddot{y}_1$ by using the constraint equation

$$y_1 + y_2 - l \;=\; 0 \tag{5.170}$$

to eliminate $y_2$ from the problem.

**5.9.5. Motion on a paraboloid revisited.** Consider a particle of mass $m$ that moves, without friction, on a smooth paraboloid while experiencing the constant force of gravity. Take

$$g(r,\theta,z) \;=\; z - r^2 \;=\; 0 \tag{5.171}$$

as your constraint equation. Write the Lagrangian for this system using cylindrical coordinates $(r,\theta,z)$ as your generalized coordinates. Augment your Lagrangian with your constraint equation and determine your equations of motion. Use your constraint equation to determine the Lagrange multiplier and to eliminate the multiplier from your equations of motion.

**5.9.6. Geodesics on a cylinder revisited.** The equation

$$g(x,y,z) \;=\; x^2 + y^2 - 1 \;=\; 0 \tag{5.172}$$

defines a right circular cylinder. Use Lagrange multipliers to show that the geodesics on the cylinder are helices.

Chapter 6

# The Second Variation

## 6.1. Introduction

In the last two chapters, we focused on generalizing the Euler–Lagrange equation to functionals with higher derivatives, multiple dependent variables, two independent variables, or constraints. Let us now return to the simple problem of minimizing or maximizing the functional

$$J[y] \;=\; \int\limits_a^b f(x, y(x), y'(x))\; dx \qquad (6.1)$$

with the boundary conditions

$$y(a) \;=\; y_a\,, \quad y(b) \;=\; y_b\,. \qquad (6.2)$$

How do we know whether we are dealing with a maximum or a minimum? Also, are there other necessary conditions besides the Euler–Lagrange equation?

Suppose that the function $y = \hat{y}(x)$, $\hat{y}(x) \in C^1[a, b]$, solves our problem and that $h(x)$ is a small deviation or *variation* from this idealized solution,

$$y(x) \;=\; \hat{y}(x) + h(x)\,, \qquad (6.3)$$

that satisfies

$$h(a) \;=\; 0 \;\text{ and }\; h(b) \;=\; 0 \qquad (6.4)$$

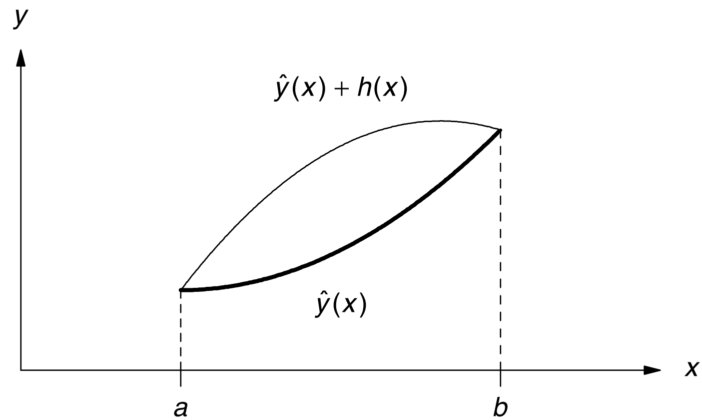



**Figure 6.1.** A small variation

(see Figure 6.1). For the time being, we will continue to use the strong norm. That is, we will restrict our attention to weak variations,

$$h(x) \;=\; \epsilon\,\eta(x)\,, \tag{6.5}$$

$\eta \in C^1[a,b]$, that satisfy the boundary conditions

$$\eta(a) \;=\; 0\,, \quad \eta(b) \;=\; 0\,. \tag{6.6}$$

By assumption, $\eta(x)$ and $\eta'(x)$ are of the same order of smallness. That is, the function $\eta(x)$ is assumed to be independent of $\epsilon$ so that, as $\epsilon$ goes to zero, the variation $h(x)$ tends to zero in both ordinate and slope.

Let us now take another look at the *total variation*

$$\Delta J \;\equiv\; J[y] - J[\hat{y}] \;=\; J[\hat{y}+h] - J[\hat{y}]\,. \tag{6.7}$$

For our simple functional, equation (6.1), and weak variation (6.5),

$$\begin{aligned}\Delta J \;&=\; \int_a^b f(x,\hat{y}+\epsilon\eta,\hat{y}'+\epsilon\eta')\;dx - \int_a^b f(x,\hat{y},\hat{y}')\;dx \qquad (6.8)\\ &=\; \int_a^b [f(x,\hat{y}+\epsilon\eta,\hat{y}'+\epsilon\eta') - f(x,\hat{y},\hat{y}')]\;dx\,.\end{aligned}$$

You will remember that we expanded the total variation in a Taylor series in $\epsilon$ and obtained

$$\Delta J \;=\; \delta J + \frac{1}{2}\,\delta^2 J + O(\epsilon^3)\,, \tag{6.9}$$

where we called

$$\delta J \;=\; \epsilon \int_a^b \left[f_y(x,\hat{y},\hat{y}')\,\eta + f_{y'}(x,\hat{y},\hat{y}')\,\eta'\right]\,dx \tag{6.10}$$

the *first variation* and

$$\delta^2 J = \epsilon^2 \int_a^b \left[f_{yy}(x,\hat{y},\hat{y}')\eta^2 + 2\,f_{yy'}(x,\hat{y},\hat{y}')\eta\eta' + f_{y'y'}(x,\hat{y},\hat{y}')\eta'^2\right]\,dx \tag{6.11}$$

the *second variation.*

We have already imposed the condition that

$$\delta J \;=\; 0\,. \tag{6.12}$$

It now follows that

$$\Delta J \;=\; \frac{1}{2}\,\delta^2 J + O(\epsilon^3)\,. \tag{6.13}$$

For sufficiently small $\epsilon$, the total variation is dominated by the second variation. It thus follows that:

***Second-variation condition:***

For the functional $J[y]$ to have a relative minimum (maximum) at $y = \hat{y}(x)$, $\hat{y}(x) \in C^1[a,b]$, it is necessary that the second variation be positive (negative) or zero,

$$\delta^2 J \;\geq\; 0 \;\;(\leq\; 0)\,, \tag{6.14}$$

for all weak $\eta(x) \in C^1[a,b]$ that vanish at $a$ and $b$.

Note that we have assumed that our extremals are continuously differentiable and that our variations are continuously differentiable and weak. These assumptions can and have been weakened. We are, however, following the historical order of events: the second variation was investigated before scientists thought about solutions with corners or about strong variations.

## 6.2. Legendre's condition

For convenience, rewrite the second variation as

$$\delta^2 J \;=\; \epsilon^2 \int_a^b (P\,\eta^2 + 2\,Q\,\eta\eta' + R\,\eta'^2)\;dx\,, \tag{6.15}$$

where

$$P \;\equiv\; f_{yy}(x,\hat{y},\hat{y}')\,,\quad Q \;\equiv\; f_{yy'}\,(x,\hat{y},\hat{y}')\,,\quad R \;=\; f_{y'y'}(x,\hat{y},\hat{y}')\,. \tag{6.16}$$

To make progress, we will transform the second variation into a more convenient form. There are several possible approaches. For historical reasons, we will follow Legendre (1788). *Warning*: There are problems with this approach. These were pointed out, early on, by Lagrange (1797). We will proceed carefully and consider both the benefits of and the problems with Legendre's approach.

For any $w \in C^1[a,b]$,

$$\int_a^b \frac{d}{dx}[w(x)\,\eta^2(x)]\;dx \;=\; w(x)\,\eta^2(x)\Big|_{x\,=\,a}^{x\,=\,b} \;=\; 0 \tag{6.17}$$

since the variation $\eta(x)$ vanishes at $a$ and $b$. If we expand the integrand and add it to the second variation, we obtain

$$\delta^2 J \;=\; \epsilon^2 \int_a^b [(P+w')\,\eta^2 + 2\,(Q+w)\,\eta\,\eta' + R\,\eta'^2]\;dx\,. \tag{6.18}$$

We may now complete the square within the integrand:

$$\delta^2 J = \epsilon^2 \int_a^b \left\{ R\left(\eta' + \frac{Q+w}{R}\,\eta\right)^2 + \left[(P+w') - \frac{(Q+w)^2}{R}\right]\eta^2 \right\}\,dx\,. \tag{6.19}$$

Up until now, we have not put any restriction on $w(x)$. Let us now impose the obvious restriction

$$w' \;=\; -P + \frac{(Q+w)^2}{R} \tag{6.20}$$

so that

$$\delta^2 J \;=\; \epsilon^2 \int_a^b R\left(\eta' + \frac{Q+w}{R}\,\eta\right)^2\, dx\,. \tag{6.21}$$

To the extent that

$$\left(\eta' + \frac{Q+w}{R}\,\eta\right)^2 \tag{6.22}$$

is nonnegative, it *appears* that the sign of the second variation is determined by the sign of $R$. Legendre claimed that $R$ must not change sign in $[a,b]$ if we are to have a relative extremum. This necessary condition is now known as *Legendre's condition.* Legendre also claimed that if $R$ is nonzero and of one sign (the *strengthened Legendre condition*), $\delta^2 J$ is of the same sign as $R$ (a sufficient condition). Legendre was correct about his necessary condition, but he was mistaken about his sufficient condition (*Legendre's fallacy*). The defect in Legendre's logic was pointed out by Lagrange (1797): Legendre's transformation assumes that differential equation (6.20) has a solution $w$ that exists on the entire interval $[a,b]$. There are, of course, existence theorems for ordinary differential equations, but for a nonlinear differential equation such as this, they only guarantee a solution locally, not on an entire interval.

**Example 6.1.**

Let

$$P \;=\; -1, \quad Q \;=\; 0, \quad R \;=\; 1, \tag{6.23}$$

so that

$$w' \;=\; 1 + w^2 \tag{6.24}$$

and let us take the initial condition

$$w(0) \;=\; 0\,. \tag{6.25}$$

Since the right-hand side of equation (6.24) is continuous as a function of $x$ and $w$ and has a continuous partial derivative with respect to $w$ in the vicinity of $x = 0$ and $w = 0$, we can expect a unique solution in some neighborhood of $(0,0)$. The function

$$w(x) \;=\; \tan x \tag{6.26}$$

is this solution. This solution fails to exist at $x = \pm\pi/2$. This is our first hint that the length of the interval $[a, b]$ may make a difference.

Let us salvage what we can before moving on:

***Legendre's condition:***

A necessary condition for the functional $J[y]$ to have a relative minimum (maximum) at $y = \hat{y}(x)$ is that

$$R(x) \equiv \frac{\partial^2 f}{\partial y'^2}(x, \hat{y}(x), \hat{y}'(x)) \geq 0 \ (\leq 0) \tag{6.27}$$

in $[a, b]$.

In the following proof and in the remainder of this chapter, we will assume, for convenience, that we are talking about minima. Feel free to work through the corresponding arguments for maxima.

**Proof.** Suppose, instead, that

$$R(c) < 0 \tag{6.28}$$

for some value $c$ in the interval $(a, b)$. Since $R(x)$ is continuous, there is a closed interval, $[c - \delta_1, c + \delta_1]$, within $[a, b]$, where $R(x) < 0$. It also follows that there is another closed interval, $[c - \delta_2, c + \delta_2]$, within $[a, b]$, where the differential equation

$$w' = -P + \frac{(Q + w)^2}{R} \tag{6.29}$$

has a continuously differentiable solution $w(x)$. Let $[x_1, x_2]$ be the smaller of these two intervals and choose $\eta(x)$ so that

$$\eta(x) = 0, \quad x \notin (x_1, x_2), \tag{6.30}$$

$$\eta(x) \neq 0, \quad x \in (x_1, x_2). \tag{6.31}$$

Since $\eta \in C^1[a, b]$,

$$\eta(x_1) = \eta(x_2) = 0 \ \text{ and } \ \eta'(x_1) = \eta'(x_2) = 0. \tag{6.32}$$

For this particular $\eta(x)$,

$$\delta^2 J \;=\; \epsilon^2 \int_{x_1}^{x_2} (P\,\eta^2 + 2\,Q\,\eta\eta' + R\,\eta'^2)\; dx\,. \tag{6.33}$$

For this interval of integration, we may safely apply Legendre's transformation, so that

$$\delta^2 J \;=\; \epsilon^2 \int_{x_1}^{x_2} R \left(\eta' + \frac{Q+w}{R}\,\eta\right)^2 \; dx\,, \tag{6.34}$$

which is clearly nonpositive.

If equation (6.34) were to vanish, we would require

$$\eta'(x) + \frac{Q+w}{R}\,\eta(x) \;\equiv\; 0 \tag{6.35}$$

for $x_1 \le x \le x_2$ since $R(x) < 0$ on this interval. This linear, first-order differential equation and the boundary condition $\eta(x_1) = 0$ would then imply that $\eta(x) \equiv 0$ for $x_1 < x < x_2$. Since we have chosen $\eta(x)$ so that it is nonzero on this open interval, we may safely conclude that the integral on the right-hand side of equation (6.34) is negative.

We have shown that if $R(c)$ is negative, we can come up with an arbitrarily small variation that causes

$$\delta^2 J \;<\; 0\,. \tag{6.36}$$

This is incompatible with the existence of a minimum. We therefore require

$$R(x) \;\ge\; 0 \tag{6.37}$$

in $[a, b]$. ♣

There are exceptional cases in which $R(x)$ has isolated zeros in the interval $[a, b]$. In most cases of interest, $R(x) > 0$ or $R(x) < 0$ everywhere in the interval $[a, b]$.

**Example 6.2** (Geodesics in the plane).

Let

$$J[y] \;=\; \int_a^b \sqrt{1+y'^2}\;dx\,. \tag{6.38}$$

Since

$$\frac{\partial^2 f}{\partial y'^2} \;=\; \frac{1}{(1+y'^2)^{3/2}} \;=\; \frac{1}{(1+m^2)^{3/2}} \tag{6.39}$$

for lines of slope $m$, it follows that

$$\left.\frac{\partial^2 f}{\partial y'^2}\right|_{\hat{y}(x)} \;>\; 0 \tag{6.40}$$

and that our extremals are not maxima but may still be minima.

**Example 6.3** (Minimal surface of revolution).

For

$$J[y] \;=\; \int_a^b y\sqrt{1+y'^2}dx\;, \tag{6.41}$$

we have

$$\frac{\partial^2 f}{\partial y'^2} \;=\; \frac{y}{(1+y'^2)^{3/2}}\,. \tag{6.42}$$

This is positive if $y(x) > 0$. All of our extremals — there may be two, one, or zero solutions for each set of boundary conditions — are potential minima.

**Example 6.4** (Geodesics on a sphere).

For

$$J[\phi] \;=\; R\int_{\theta_1}^{\theta_2} \sqrt{1+\sin^2\theta\left(\frac{d\phi}{d\theta}\right)^2}\;d\theta\,, \tag{6.43}$$

we have

$$\frac{\partial^2 f}{\partial \phi'^2} \;=\; \frac{\sin^2\theta}{[1+\sin^2\theta\,(\phi')^2]^{3/2}} \;>\; 0 \tag{6.44}$$

so that arcs of great circles are, potentially, minima.

## 6.3. Jacobi's condition

Legendre attempted to show that the strengthened Legendre condition is a sufficient condition for relative minimum. His reasoning was flawed. He had, as pointed out by Lagrange, implicitly assumed that equation (6.20) has a solution that is finite and continuous over the entire interval $[a, b]$. This may not be true. It is clear, therefore, that we need to study differential equation (6.20) in greater detail. In the course of studying this equation, we will uncover new (necessary and sufficient) conditions for a weak relative minimum.

Differential equation (6.20),

$$w' = -P + \frac{(Q+w)^2}{R}, \tag{6.45}$$

is a *Riccati equation.* You may not remember much about Riccati equations. So here are some basic facts about these equations. A Riccati equation is any differential equation of the general form

$$y' = a(x)y^2 + b(x)y + c(x). \tag{6.46}$$

If $a(x) = 0$, Riccati's equation reduces to a linear equation; for $c(x) = 0$, we get a Bernoulli equation.

In most cases, we cannot solve Riccati equations in closed form. If, however, we know one particular solution, the general solution *can* be calculated. Indeed, let $y_p(x)$ be a particular solution to Riccati equation (6.46),

$$y_p' = ay_p^2 + by_p + c. \tag{6.47}$$

Introduce a new variable $z(x)$ that measures the difference between the general solution and the particular solution,

$$z(x) \equiv y(x) - y_p(x). \tag{6.48}$$

It quickly follows that

$$z' = az^2 + (2ay_p + b)z. \tag{6.49}$$

This last equation is a Bernoulli equation; it can be solved exactly.

**6.3.1. The Jacobi equation.** One of the reasons that Riccati equations are important is that every linear, second-order, homogeneous, ordinary differential equation can be turned into a Riccati equation, and vice versa. To turn equation (6.45) into a linear, second-order differential equation, we introduce

$$w(x) \;=\; -Q - R\,\frac{u'(x)}{u(x)}\,, \tag{6.50}$$

under the assumption that

$$u(x) \;\neq\; 0\,, \quad x \in [a,b]\,. \tag{6.51}$$

The resulting equation,

$$\frac{d}{dx}(Ru') + (Q' - P)\,u \;=\; 0\,, \tag{6.52}$$

is known as the *Jacobi equation,* after Carl Gustav Jacob Jacobi (1804–1851). It may also be rewritten as

$$\frac{d^2u}{dx^2} + \frac{R'}{R}\frac{du}{dx} + \frac{(Q' - P)}{R}\,u \;=\; 0\,. \tag{6.53}$$

In terms of $u$, the second variation,

$$\delta^2 J \;=\; \epsilon^2 \int_a^b R\left(\eta' + \frac{Q + w}{R}\,\eta\right)^2 dx\,, \tag{6.54}$$

now simplifies to

$$\delta^2 J \;=\; \epsilon^2 \int_a^b R\left(\eta' - \frac{u'}{u}\,\eta\right)^2 dx\,. \tag{6.55}$$

Having introduced the Jacobi equation and having transformed the second variation using solutions of the Jacobi equation, we can now state a simple sufficiency condition that guarantees that the second variation is positive definite:

***A weak sufficiency condition:***

For $y = \hat{y}(x)$, if

(a) $R(x) > 0$, for all $x \in [a, b]$, and
(b) the Jacobi equation has a solution $u = u(x) \neq 0$, for all $x \in [a, b]$,

then $\delta^2 J$ is positive definite. That is, $\delta^2 J > 0$ for every admissible weak variation $\eta(x)$ that is not identically zero.

**Proof.** If the Jacobi equation has a solution that does not vanish anywhere on the entire interval $[a, b]$, we can carry out Legendre's transformation of the second variation, as described above. For $R(x) > 0$, the second variation can vanish only if

$$\eta' - \frac{u'}{u}\eta \;\equiv\; 0 \tag{6.56}$$

for all $x \in [a, b]$. Since

$$\eta(a) \;=\; 0\,, \quad \eta(b) \;=\; 0\,, \tag{6.57}$$

and since $u(x)$ does not vanish at these endpoints, the only solution to this linear, first-order differential equation is

$$\eta(x) \;=\; 0\,. \tag{6.58}$$

♣

If *every* solution of the Jacobi equation vanishes for at least one point of $[a, b]$, we will not be able to carry out Legendre's transformation of the second variation over the whole interval. At the very least, we will then lose our sufficiency condition for the positive definiteness of the second variation. Worse yet, if there are two points of $[a, b)$ where a nontrivial solution of the Jacobi equation vanishes, the second variation can, in general, be made negative. This will ultimately lead to a new necessary condition for a weak relative minimum.

We have gone through a number of gyrations to derive the Jacobi equation. We will shortly solve Jacobi's equation. Before doing so, however, let's take a brief, seemingly unrelated, detour and see yet another way in which this equation arises. This detour will also provide us an alternative way of writing the second variation that will prove useful later in this chapter.

Accordingly, let us write the integrand of the second variation as a function of $\eta$ and $\eta'$,

$$\begin{aligned} \delta^2 J &= \epsilon^2 \int_a^b (P\,\eta^2 + 2\,Q\,\eta\,\eta' + R\,{\eta'}^2)\;dx \\ &= \epsilon^2 \int_a^b 2\,\Omega(\eta, \eta')\;dx\,. \end{aligned} \tag{6.59}$$

Since the integrand is homogeneous of degree two in $\eta$ and $\eta'$, it quickly follows that

$$2\,\Omega \;=\; \eta\,\frac{\partial\Omega}{\partial\eta} + \eta'\,\frac{\partial\Omega}{\partial\eta'}\,. \tag{6.60}$$

(You may recognize this statement as a simple example of Euler's identity for homogeneous functions. We will discuss this identity, more fully, in Chapter 8.)

We may now write the second variation as

$$\delta^2 J \;=\; \epsilon^2 \int_a^b \left(\eta\,\frac{\partial\Omega}{\partial\eta} + \eta'\,\frac{\partial\Omega}{\partial\eta'}\right)\;dx\,. \tag{6.61}$$

If we integrate the second term in the integral by parts, using the fact that $\eta(x)$ vanishes at the endpoints, we obtain

$$\delta^2 J \;=\; \epsilon^2 \int_a^b \eta\left[\frac{\partial\Omega}{\partial\eta} - \frac{d}{dx}\left(\frac{\partial\Omega}{\partial\eta'}\right)\right]\;dx\,. \tag{6.62}$$

Interestingly,

$$\frac{\partial\Omega}{\partial\eta} - \frac{d}{dx}\left(\frac{\partial\Omega}{\partial\eta'}\right) \;=\; (P - Q')\,\eta - \frac{d}{dx}(R\,\eta') \;\equiv\; \Psi(\eta) \tag{6.63}$$

so that

$$\delta^2 J \;=\; \epsilon^2 \int_a^b \eta\,\Psi(\eta)\;dx\,. \tag{6.64}$$

$\Psi(\eta)$ is just the negative of the Jacobi equation (with $\eta(x)$ as the dependent variable). It is often said that the Jacobi equation is the Euler–Lagrange equation of the secondary or accessory variational problem. Keep equation (6.64) in mind. We will need this equation shortly.

**6.3.2. Solutions of the Jacobi equation.** It is clear that we must solve the Jacobi equation. Fortunately, Jacobi (1837) invented an amazing method of solution. He showed that the general solution of the Jacobi equation can be derived by straightforward differentiation of the general solution of the Euler–Lagrange equation.

Suppose that we have the general solution,

$$y = \hat{y}(x, \alpha, \beta)\,, \tag{6.65}$$

to the Euler–Lagrange equation that arises out of the first variation. Since the Euler–Lagrange equation is a second-order differential equation, this general solution contains the two constants of integration $\alpha$ and $\beta$.

If we insert this general solution into the Euler–Lagrange equation, we obtain

$$\begin{aligned} &\frac{\partial f}{\partial y}\,(x, \hat{y}(x,\alpha,\beta), \hat{y}'(x,\alpha,\beta)) \\ &\qquad - \frac{d}{dx}\left[\frac{\partial f}{\partial y'}\,(x, \hat{y}(x,\alpha,\beta), \hat{y}'(x,\alpha,\beta))\right] = 0\,. \end{aligned} \tag{6.66}$$

This identity is satisfied for all values of $x$, $\alpha$, and $\beta$, and we are therefore free to differentiate this identity with respect to either $\alpha$ or $\beta$. We will do so for $\alpha$, taking the additional liberty of reversing the order of differentiation between $x$ and $\alpha$:

$$\frac{\partial^2 f}{\partial y^2}\frac{\partial \hat{y}}{\partial \alpha} + \frac{\partial^2 f}{\partial y\,\partial y'}\frac{\partial \hat{y}'}{\partial \alpha} - \frac{d}{dx}\left(\frac{\partial^2 f}{\partial y'\,\partial y}\frac{\partial \hat{y}}{\partial \alpha} + \frac{\partial^2 f}{\partial y'^2}\frac{\partial \hat{y}'}{\partial \alpha}\right) = 0\,. \tag{6.67}$$

We may now rewrite this as

$$P\,\frac{\partial \hat{y}}{\partial \alpha} + Q\,\frac{\partial \hat{y}'}{\partial \alpha} - \frac{d}{dx}\left(Q\,\frac{\partial \hat{y}}{\partial \alpha} + R\,\frac{\partial \hat{y}'}{\partial \alpha}\right) = 0 \tag{6.68}$$

or

$$(P - Q')\,\frac{\partial \hat{y}}{\partial \alpha} + (Q - Q)\,\frac{\partial \hat{y}'}{\partial \alpha} - \frac{d}{dx}\left(R\,\frac{\partial \hat{y}'}{\partial \alpha}\right) = 0\,. \tag{6.69}$$

This simplifies to

$$(P - Q')\,\frac{\partial \hat{y}}{\partial \alpha} - \frac{d}{dx}\left(R\,\frac{\partial \hat{y}'}{\partial \alpha}\right) \;=\; 0\,. \tag{6.70}$$

But this is just the Jacobi equation with

$$u \;=\; \frac{\partial \hat{y}}{\partial \alpha} \quad\text{and}\quad u' \;=\; \frac{\partial \hat{y}'}{\partial \alpha}\,. \tag{6.71}$$

We could just as easily have differentiated with respect to $\beta$. It appears that we have proven the following important theorem:

***Jacobi's theorem:***

If $y = \hat{y}(x, \alpha, \beta)$ is the general solution of an Euler–Lagrange equation, then the corresponding Jacobi equation,

$$\frac{d}{dx}(Ru') \;+\; (Q' \;-\; P)u \;=\; 0\,, \tag{6.72}$$

has the two particular solutions

$$u_1(x) \;=\; \frac{\partial \hat{y}}{\partial \alpha} \quad\text{and}\quad u_2(x) \;=\; \frac{\partial \hat{y}}{\partial \beta}\,. \tag{6.73}$$

These two solutions are, in general, linearly independent. The general solution of Jacobi's differential equation is, typically,

$$u(x) \;=\; c_1\,u_1(x) + c_2\,u_2(x)\,. \tag{6.74}$$

**6.3.3. Jacobi's criterion.** Let us now look at one particularly convenient solution of Jacobi's equation:

$$\Delta(x, a) \;=\; u_2(a)\,u_1(x) - u_1(a)\,u_2(x)\,. \tag{6.75}$$

This solution has been chosen so that it vanishes at $x = a$.

If the next zero of $\Delta(x, a)$, $x = \overline{a}$, occurs past $x = b$ (see Figure 6.2), we can safely construct a solution $u(x)$ that does not vanish on the closed interval $[a, b]$ by considering $\Delta(x, a - \epsilon)$ for sufficiently small $\epsilon$. We are then guaranteed, by our weak sufficiency condition, that $\delta^2 J > 0$ for all admissible and nonvanishing $\eta(x)$.

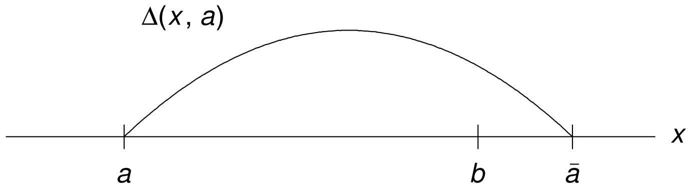



**Figure 6.2.** $\overline{a} > b$

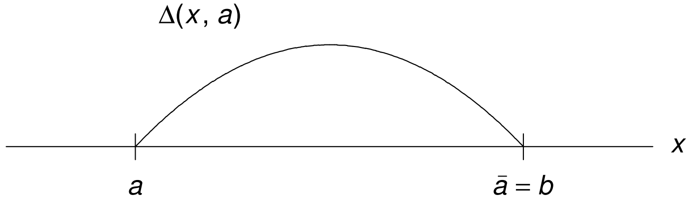



**Figure 6.3.** $\overline{a} = b$

If the next zero, $x = \overline{a}$, occurs at $x = b$ (see Figure 6.3), the situation changes dramatically. We can now find a nontrivial, admissible variation that causes the second variation to vanish. To see this, first note that restriction (6.51) implies that we can no longer write the second variation as

$$\delta^2 J \;=\; \epsilon^2 \int_a^b R\left(\eta' - \frac{u'}{u}\,\eta\right)^2 dx\,. \tag{6.76}$$

We can, however, still use equation (6.64) to write the second variation as

$$\delta^2 J \;=\; \epsilon^2 \int_a^b \eta\,\Psi(\eta)\;dx\,. \tag{6.77}$$

If we now choose $\eta(x)$ to be a multiple of $u(x) = \Delta(x, a)$, $\eta(x)$ satisfies the boundary conditions $\eta(a) = \eta(b) = 0$. For this choice, moreover, $\Psi(\eta) \equiv 0$ so that the second variation vanishes.

Once the second variation vanishes for nontrivial $\eta(x)$, the sign of the total variation, $\Delta J$, depends on the sign of the third variation, $\delta^3 J$, which is of odd power in $\epsilon$. In general, the third variation can be made negative by choosing the sign of $\epsilon$ appropriately. If the third variation should (miraculously) be equal to zero, we would need to look at the fourth variation.

Jacobi asserted, but did not prove, that there will be *no* minimum if the next zero of $\Delta(x, a)$, at $x = \overline{a}$, occurs before $x = b$. In 1855, J. Bertrand conjectured that Jacobi may have overstated his results and that while $b < \overline{a}$ may be useful as part of a sufficiency condition, it may not be necessary. In the same year, however, Ossian Bonnet showed, in the context of geodesics, that the second variation can be made negative if $\overline{a} < b$. See Todhunter (2005) for details. The earliest general proof of the necessity of Jacobi's condition is due to Erdmann (1878).

Questions about the necessity of Jacobi's condition arose late enough that most proofs use varied curves with corners (which we have, so far, ignored) instead of weak, continuously differentiable variations. One of the few proofs that does work for weak, continuously differentiable variations is the proof of Gelfand and Fomin (1963) and van Brunt (2004), which relies on a homotopy argument.

Basically, these authors assume a one-parameter family of *positive-definite* functionals of the form

$$K(\mu) \;=\; \mu\,\delta^2 J + (1-\mu)\,\epsilon^2 \int_a^b \eta'^2\;dx\,, \tag{6.78}$$

where $\mu \in [0, 1)$. For $\mu = 0$, $K(\mu)$ reduces to a simple quadratic functional and $\Delta(x, a)$ does not vanish inside $(a, b)$. (Verify this.) Gelfand and Fomin (1963) and van Brunt (2004) then prove that a zero of $\Delta(x, a)$ does not enter $(a, b)$ as one increases $\mu$ from 0 to 1. They thereby show that the only way that we can have a relative minimum is if $\Delta(x, a)$ has no zero in $(a, b)$. This is more than we want to prove in an introduction to the subject. So, let me instead refer you to the appropriate textbooks and state the main conclusion.

***Jacobi's condition:***

A necessary condition for a relative minimum (maximum) is that

$$\Delta(x, a) \neq 0 \tag{6.79}$$

for all values of $x$ in the open interval $a < x < b$.

**6.3.4. Conjugate points.** The first zero, $x = \overline{a}$, of $\Delta(x, a)$ that follows $x = a$ is so important that we say that $\overline{a}$ is *conjugate* to $a$. The point $\overline{A}$ on the extremal with abscissa $\overline{a}$ is, in turn, conjugate[1] to the point $A$ with abscissa $a$.

There are two ways to determine conjugate points.

*Analytic method:*

The value $\overline{a}$ satisfies the equation

$$\Delta(\overline{a}, a) = u_2(a)\, u_1(\overline{a}) - u_1(a)\, u_2(\overline{a}) = 0\,. \tag{6.80}$$

It follows that

$$\frac{u_1(\overline{a})}{u_2(\overline{a})} = \frac{u_1(a)}{u_2(a)} \tag{6.81}$$

and that

$$\left.\frac{\frac{\partial \hat{y}}{\partial \alpha}}{\frac{\partial \hat{y}}{\partial \beta}}\right|_{x=\overline{a}} = \left.\frac{\frac{\partial \hat{y}}{\partial \alpha}}{\frac{\partial \hat{y}}{\partial \beta}}\right|_{x=a}\,. \tag{6.82}$$

Here, we have used Jacobi's result that one can obtain particular solutions of Jacobi's equation by differentiating a two-parameter general solution of the corresponding Euler–Lagrange equation with respect to the parameters. The last displayed equation can, in some instances, be solved for $x = \overline{a}$.

*Geometrical method:*

Let $u(x)$ be a solution of the Jacobi equation, let $\hat{y}(x, \alpha, \beta)$ be the general solution of the corresponding Euler–Lagrange equation, and let $\gamma$ be a constant. Consider the two curves

$$y = \hat{y}(x, \alpha, \beta) \quad \text{and} \quad y = \hat{y}(x, \alpha, \beta) + \gamma\, u(x)\,. \tag{6.83}$$

[1] Many investigators refer to both $\overline{a}$ and $\overline{A}$ as conjugate points while other investigators distinguish between the conjugate value $\overline{a}$ and the conjugate point $\overline{A}$. Similarly, some investigators call any zero of $\Delta(x, a)$ that follows $x = \overline{a}$ a conjugate point while other investigators reserve this term for the first such zero.

These two curves intersect whenever

$$u(x) \;=\; 0\,. \tag{6.84}$$

If these two curves intersect at $A$, they will also intersect at the conjugate point, $\overline{A}$.

Now, consider two nearby extremals,

$$y \;=\; \hat{y}(x, \alpha, \beta) \tag{6.85}$$

and

$$y \;=\; \hat{y}(x, \alpha + \Delta\alpha, \beta)\,, \tag{6.86}$$

for $\Delta\alpha$ small. Ignoring higher-order terms, the second extremal may be approximated by

$$y \;=\; \hat{y}(x, \alpha, \beta) + \frac{\partial \hat{y}}{\partial \alpha}\,\Delta\alpha \tag{6.87}$$

or

$$y \;=\; \hat{y}(x, \alpha, \beta) + \Delta\alpha \cdot u_1(x)\,. \tag{6.88}$$

It now stands to reason that if neighboring extremals (6.85) and (6.86) intersect at $A$, they will also intersect at or, because of the approximation, near the conjugate point $\overline{A}$.

Let's make all of this a little more precise. Let us reduce our two-parameter family of extremals to a one-parameter family of extremals,

$$y \;=\; \hat{y}(x, c)\,, \tag{6.89}$$

by restricting our attention to a pencil of extremals emanating out of the fixed point $A = (a, y_a)$ (see Figure 6.4). Let us now consider two "neighboring" curves,

$$y = \hat{y}(x, c) \;\text{ and }\; y = \hat{y}(x, c + \Delta c)\,, \tag{6.90}$$

of this one-parameter family.

These two extremals intersect if they share the same coordinates. We can use the equation

$$\hat{y}(x, c + \Delta c) \;=\; \hat{y}(x, c) \tag{6.91}$$

to determine the abscissa of the point of intersection and the equation

$$y \;=\; \hat{y}(x, c) \tag{6.92}$$

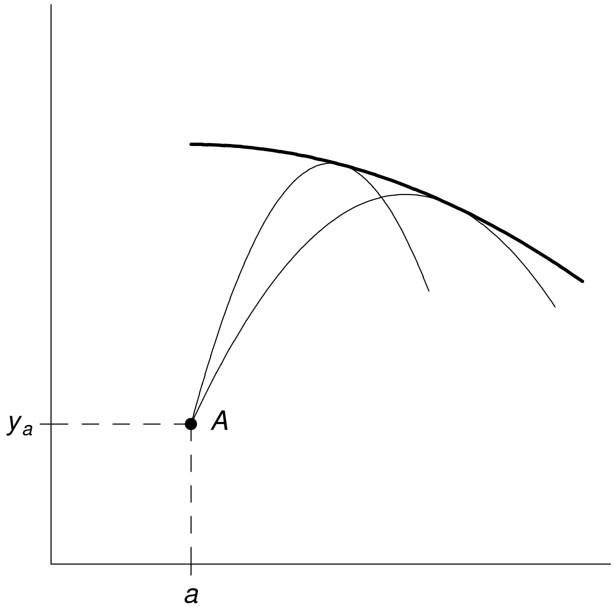



**Figure 6.4.** Envelope

to then determine the ordinate. We may also rewrite these conditions as

$$y = \hat{y}(x, c) \tag{6.93}$$

and

$$\frac{\hat{y}(x, c + \Delta c) - \hat{y}(x, c)}{\Delta c} = 0\,. \tag{6.94}$$

Here, we have divided by $\Delta c$ so that we can get a nontrivial result in the limit as $\Delta c$ goes to zero.

In that limit,

$$y = \hat{y}(x, c)\,, \quad \frac{\partial \hat{y}}{\partial c}(x, c) = 0\,, \tag{6.95}$$

we capture our conjugate point. But, this is also the *c-discriminant,* the set of equations that specifies the *envelope* of our one-parameter family. Thus, if a family of extremals, emanating from a point $A$, has an envelope, the conjugate point for a member of this family is the point of contact of that extremal with the envelope.

**Example 6.5.**

Consider the functional

$$J[y] \;=\; \int_0^b (y'^2 - y^2)\;dx \tag{6.96}$$

with $0 < b < \pi$ and the boundary conditions

$$y(0) \;=\; 0, \quad y(b) \;=\; 1\,. \tag{6.97}$$

The Euler–Lagrange equation for this problem is just

$$y'' + y \;=\; 0\,. \tag{6.98}$$

The general, two-parameter, solution to this equation is

$$\hat{y}(x,\alpha,\beta) \;=\; \alpha\;\sin\,x + \beta\;\cos\,x\,. \tag{6.99}$$

Upon applying the boundary conditions, we discover that

$$\alpha \;=\; \frac{1}{\sin\,b} \quad \text{and} \quad \beta \;=\; 0\,. \tag{6.100}$$

This leaves us with the extremal

$$\hat{y}(x) \;=\; \frac{\sin\,x}{\sin\,b} \tag{6.101}$$

(see Figure 6.5).

Upon checking the Legendre condition, we discover that

$$\frac{\partial^2 f}{\partial y'^2} \;=\; 2 \;>\; 0\,. \tag{6.102}$$

The strengthened Legendre condition is satisfied.

Since

$$P \;=\; f_{yy}(x,\hat{y},\hat{y}') \;=\; -2\,, \tag{6.103}$$

$$Q \;=\; f_{yy'}(x,\hat{y},\hat{y}') \;=\; 0\,, \tag{6.104}$$

$$R \;=\; f_{y'y'}(x,\hat{y},\hat{y}') \;=\; 2\,, \tag{6.105}$$

Jacobi's equation,

$$\frac{d}{dx}(Ru') + (Q' - P)\,u \;=\; 0\,, \tag{6.106}$$

reduces to

$$u'' + u \;=\; 0\,. \tag{6.107}$$

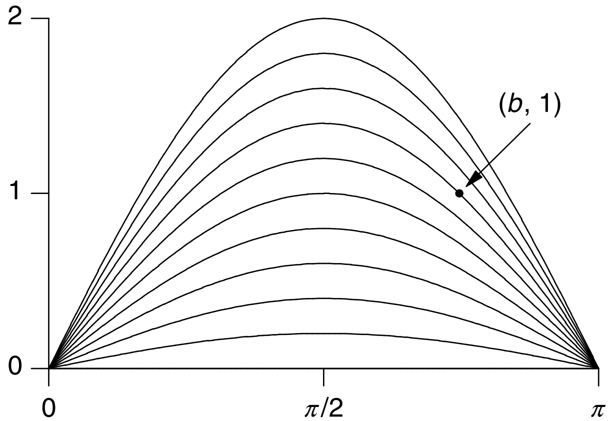



**Figure 6.5.** One-parameter family of extremals

Jacobi's direct method of solution,

$$u(x) \;=\; c_1\, u_1(x) + c_2\, u_2(x) \;=\; c_1\, \frac{\partial \hat{y}}{\partial \alpha} + c_2\, \frac{\partial \hat{y}}{\partial \beta}\,, \tag{6.108}$$

generates the obvious general solution to this equation,

$$u(x) \;=\; c_1 \sin x + c_2 \cos x\,. \tag{6.109}$$

We can now determine the conjugate value analytically. The value $\overline{a}$ satisfies

$$\frac{u_1(\overline{a})}{u_2(\overline{a})} \;=\; \frac{u_1(a)}{u_2(a)}\,, \tag{6.110}$$

which, in this case, yields

$$\frac{\sin \overline{a}}{\cos \overline{a}} \;=\; \frac{\sin 0}{\cos 0} \tag{6.111}$$

or

$$\tan \overline{a} \;=\; \tan 0 \;=\; 0\,. \tag{6.112}$$

It follows that

$$\overline{a} \;=\; \pi\,. \tag{6.113}$$

We can also determine the conjugate point graphically by looking at a one-parameter family of extremals,

$$y(x) \;=\; \alpha \sin x\,, \tag{6.114}$$

emanating out of the origin (see Figure 6.5). It is clear that the "true" extremal and its close neighbors intersect at $\overline{a} = \pi$.

## 6.4. Case study: The catenoid revisited

Let us return now to the problem of minimizing the area

$$J[y] \;=\; 2\pi \int_a^b y(x)\sqrt{1+y'^2}\; dx \tag{6.115}$$

of a surface of revolution subject to the boundary conditions

$$y(a) \;=\; y_a\,, \quad y(b) \;=\; y_b \tag{6.116}$$

in the plane. We saw, much earlier, that our general solution for this problem is the catenary

$$\hat{y}(x) \;=\; \alpha \cosh\left(\frac{x-\beta}{\alpha}\right). \tag{6.117}$$

To apply Legendre's test, we must determine the sign of

$$\frac{\partial^2 f}{\partial y'^2} \;=\; \frac{y}{(1+y'^2)^{3/2}} \tag{6.118}$$

along our solution. For

$$y(x) \;=\; \alpha \cosh\left(\frac{x-\beta}{\alpha}\right), \tag{6.119}$$

we have

$$\frac{\partial^2 f}{\partial y'^2} \;=\; \frac{\alpha}{\cosh^2\left(\frac{x-\beta}{\alpha}\right)}. \tag{6.120}$$

So, for $\alpha > 0$, we are talking about minima.

There are several ways to look for conjugate points and to check Jacobi's condition for our catenary. We can certainly take the geometric approach. If we, for example, impose the single boundary condition $y(0) = 1$, our catenary must now satisfy

$$1 \;=\; \alpha \cosh\frac{\beta}{\alpha}. \tag{6.121}$$

Let us denote $\lambda = \beta/\alpha$ so that

$$\alpha \;=\; \frac{1}{\cosh\lambda}. \tag{6.122}$$

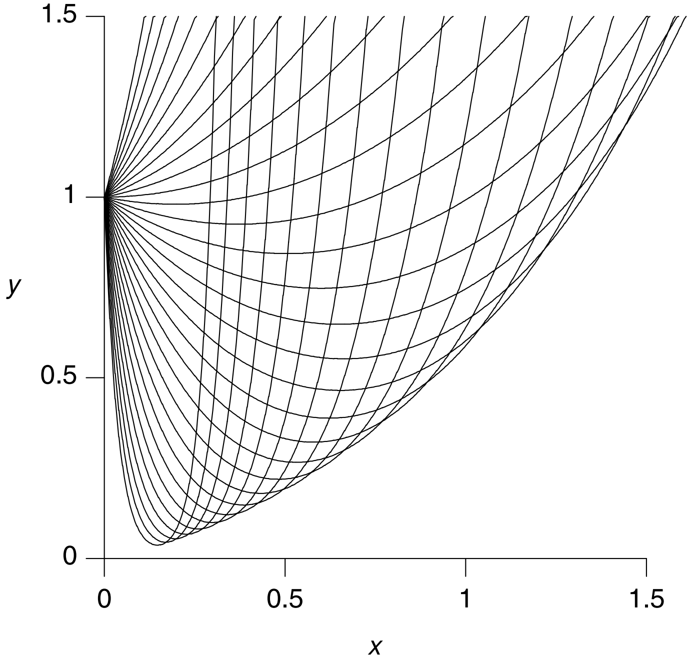



**Figure 6.6.** One-parameter family of catenaries

Our two-parameter family of catenaries now reduces to the one-parameter family

$$y(x, \lambda) \;=\; \frac{\cosh\,(x\,\cosh\,\lambda - \lambda)}{\cosh\,\lambda}\,. \tag{6.123}$$

As we vary $\lambda$ (see Figure 6.6), we obtain a one-parameter family or "pencil" of catenaries emanating from $(1, 0)$. The members of this family form an envelope that passes through the origin. Every point $P$ above this envelope is visited by two different catenaries. One of the two catenaries touches the envelope before passing on to $P$. We have seen that this catenary cannot be a minimizing curve. The other catenary, which does not touch the envelope, is a relative minimum.

We can also take an analytic approach. As a prelude, consider the *symmetric* boundary conditions

$$y(-h) \;=\; k\,, \quad y(h) \;=\; k\,. \tag{6.124}$$

These boundary conditions imply that $\beta = 0$ so that catenary (6.117) satisfies

$$\alpha \cosh \frac{-h}{\alpha} \;=\; k \;=\; \alpha \cosh \frac{h}{\alpha}\,. \tag{6.125}$$

If we let

$$z \;\equiv\; \frac{h}{\alpha}\,, \tag{6.126}$$

we may write

$$\cosh z \;=\; m\,z\,, \tag{6.127}$$

where

$$m \;\equiv\; \frac{k}{h}\,. \tag{6.128}$$

For $m = k/h$ sufficiently large, we have two roots and *two* catenaries that satisfy our boundary conditions. If our boundaries are far enough apart, there are no catenary solutions. These two extremes are separated by a critical case, $m = m_c$, $z = z_c$, that corresponds to the double root

$$\cosh z_c \;=\; m_c\, z_c\,, \quad \sinh z_c \;=\; m_c\,. \tag{6.129}$$

We can eliminate $m_c$ between these two equations to get

$$\coth z_c \;=\; z_c\,, \tag{6.130}$$

which has the solution

$$z_c \;\approx\; 1.199679\,. \tag{6.131}$$

It now follows that

$$m_c \;\approx\; \sinh(1.199679) \;\approx\; 1.508880\,. \tag{6.132}$$

For $m > m_c \approx 1.508880$, we have two roots (Figure 6.7) and two catenaries (Figure 6.8) that satisfy our boundary conditions. One catenary is shallow and one is deep. I suggested earlier that the deep solution is a spurious solution while the shallow catenary is a weak relative minimum. Let us see what our analytic conjugate-point criterion has to say about this problem.

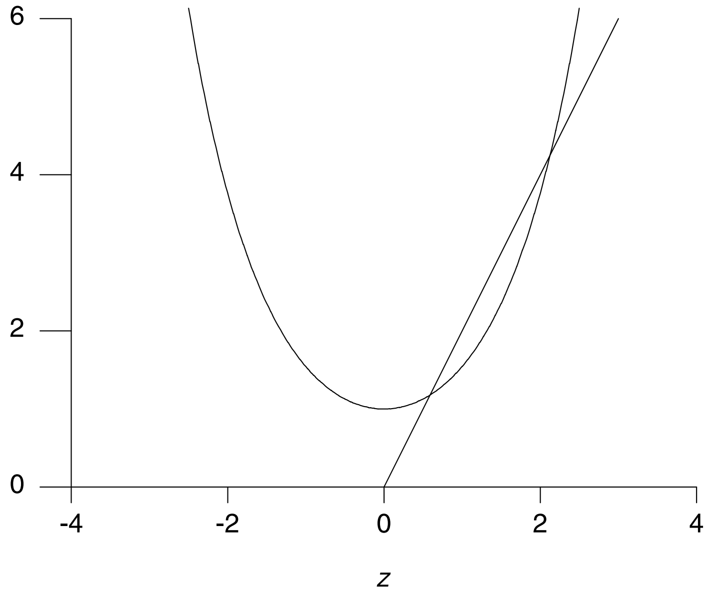



**Figure 6.7.** Two roots

The two linearly independent solutions of Jacobi's equation are

$$u_1(x) = \frac{\partial \hat{y}}{\partial \alpha} \tag{6.133}$$
$$= \cosh\left(\frac{x-\beta}{\alpha}\right) - \left(\frac{x-\beta}{\alpha}\right) \sinh\left(\frac{x-\beta}{\alpha}\right)$$

and

$$u_2(x) = \frac{\partial \hat{y}}{\partial \beta} = -\sinh\left(\frac{x-\beta}{\alpha}\right) . \tag{6.134}$$

We can now find the $\overline{a}$ that is conjugate to $x = a$ by solving the equation

$$\frac{u_1(\overline{a})}{u_2(\overline{a})} = \frac{u_1(a)}{u_2(a)} . \tag{6.135}$$

This equation reduces to

$$\overline{z} - \coth \overline{z} = z - \coth z , \tag{6.136}$$

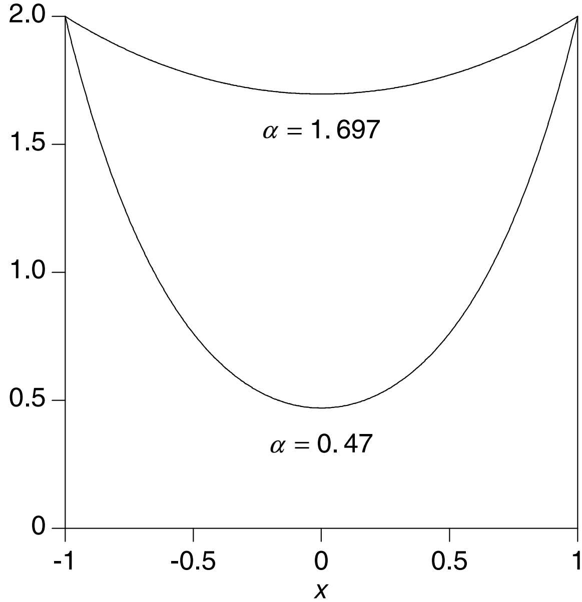



**Figure 6.8.** Two catenaries

where, now,

$$\bar{z} \equiv \frac{\bar{a}-\beta}{\alpha} \quad \text{and} \quad z \equiv \frac{a-\beta}{\alpha}\,. \tag{6.137}$$

Since the plot of $z - \coth z$ (see Figure 6.9) increases monotonically from $-\infty$ to $+\infty$ in both the left-half and the right-half planes, we see that there is one positive $\bar{z}$ for every *negative* value of $z$.

We also see that $\bar{z}$ is bigger (in magnitude) than $z$ above the $z$-axis and smaller than $z$ below the $z$-axis. The left branch of the function $z - \coth z$ intersects the $z$-axis at

$$z_c \;=\; -1.199679\,. \tag{6.138}$$

For our symmetric boundary conditions

$$y(-h) \;=\; k\,, \quad y(h) \;=\; k\,, \tag{6.139}$$

$\beta = 0$, and we now take $z$ to be negative,

$$z \;=\; \frac{a}{\alpha} \;=\; -\frac{h}{\alpha}\,. \tag{6.140}$$

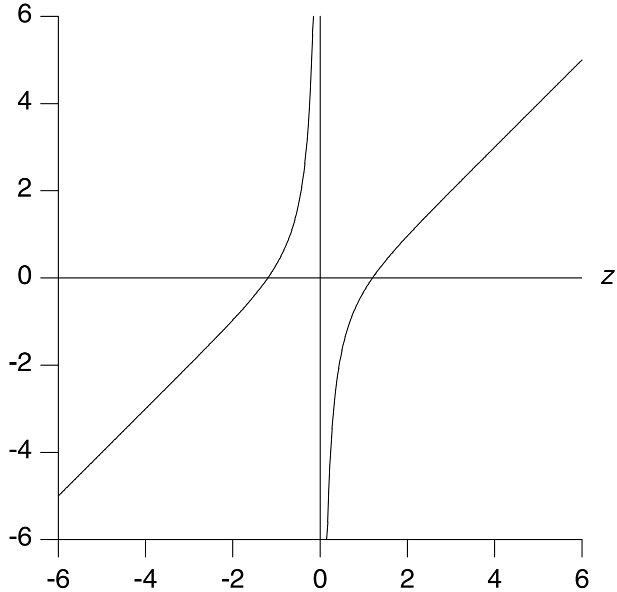



**Figure 6.9.** Plot of $z - \coth z$

For these boundary conditions, Jacobi's condition is violated for $z < z_c$ since it then follows that $|\overline{a}| < |a|$. For this symmetric problem, our high $\alpha$ solution satisfies Jacobi's condition and our low $\alpha$ solution violates Jacobi's condition.

We can also find the conjugate point $\overline{A}$ of a point $A$ on the catenary

$$\hat{y}(x) \;=\; \alpha \cosh\left(\frac{x-\beta}{\alpha}\right) \tag{6.141}$$

using a simple geometric construction, starting with our analytic criterion. In particular, Lindelöf and Moigno (1861) showed that *the tangents to the catenary at $A$ and $\overline{A}$ meet on the $x$-axis.*

To see this, consider our simple catenary (see Figure 6.10). The vertex of the catenary is $V$ and the point $U$ that lies on the $x$-axis beneath the vertex is a distance $\beta$ from the origin. Let us now pick a point $P$ on the catenary, with abscissa $x$. Let $M$ be the point directly

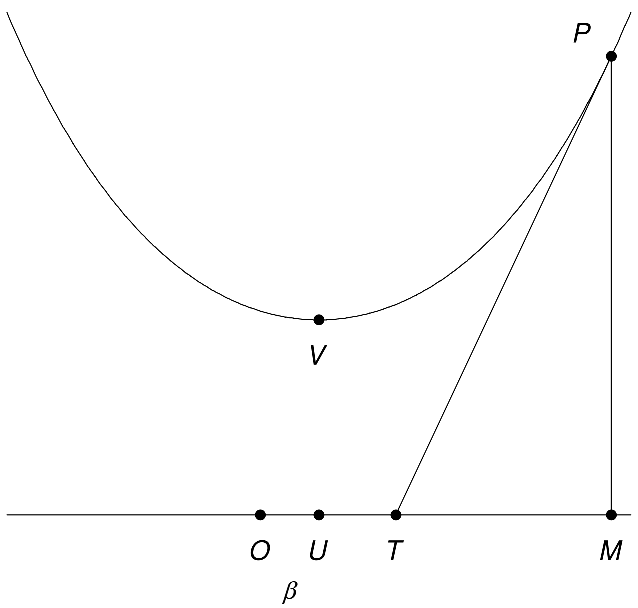



**Figure 6.10.** A simple geometric construction

underneath $P$ on the $x$-axis and let $T$ be the point of intersection of the tangent to the catenary at $P$ with the $x$-axis. It now follows that

$$TM = \frac{MP}{\tan MTP} = \frac{\alpha \cosh\left(\frac{x-\beta}{\alpha}\right)}{\sinh\left(\frac{x-\beta}{\alpha}\right)}. \tag{6.142}$$

Thus,

$$\frac{UT}{\alpha} = \frac{UM - TM}{\alpha} = \left(\frac{x-\beta}{\alpha}\right) - \coth\left(\frac{x-\beta}{\alpha}\right). \tag{6.143}$$

But, by equation (6.136), the tangents to the catenary at $A$ and $\overline{A}$ then meet at the same point on the $x$-axis.

For example, for the deep catenary shown in Figure 6.8, with $\alpha = 0.47$ and

$$z = -\frac{h}{\alpha} = -\frac{1}{\alpha} \approx -2.1276\,, \tag{6.144}$$

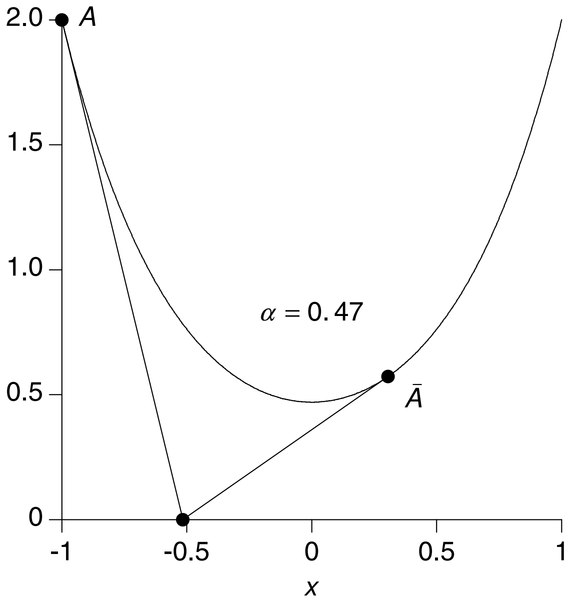



**Figure 6.11.** Conjugate point for a deep catenary

we see that

$$z \; < \; z_c \; = \; -1.199679\,. \tag{6.145}$$

Our catenary should, in other words, violate Jacobi's condition. Figure 6.11 shows that the deep catenary extends far beyond conjugate point $\overline{A}$.

## 6.5. Recommended reading

Many books on the calculus of variations discuss the second variation. See, for example, Akhiezer (1962, 1988), Bolza (1973), Brechtken-Manderscheid (1991), Forsyth (1927), Fox (1950), Gelfand and Fomin (1963), Sagan (1969), and van Brunt (2004).

Refer to Fraser (2005b) for more about the importance of Lagrange (1797) and for more about Lagrange's critique of Legendre (1788). Goldstine (1980), Kolmogorov and Yushkevich (1998), and

Todhunter (2005) review the history of the study of the second variation. See Todhunter (2005) for an English translation of Jacobi (1837).

O'Reilly and Peters (2011) describe the use of the second variation to determine the stability of elastic systems.

Morse (1934) extended Jacobi's condition by using the number of conjugate points as an index that quantifies the dimension of the space for which $\delta^2 J$ is negative. See Milnor (1963) for more on Morse theory and see the recent review article by Manning (2009) for a related perspective.

The theory of conjugate points has been extended to variational problems with isoperimetric constraints (Bolza, 1973). The notion of an index has also been extended to isoperimetric problems and has been used to study the stability of capillary surfaces (Gillette and Dyson, 1971; Lowry and Steen, 1995), DNA minicircles (Manning et al., 1998), twisted elastic struts (Hoffman et al., 2002), and twisted elastic loops (Hoffman, 2005).

## 6.6. Exercises

**6.6.1. Third variation.** Consider a functional of the form

$$J[y] = \int_a^b f(x, y(x), y'(x))\, dx\,. \tag{6.146}$$

Derive an expression for the third variation of $J[y]$.

**6.6.2. Higher-order derivatives.** Consider a functional of the form

$$J[y] = \int_a^b f(x, y(x), y'(x), y''(x))\, dx\,. \tag{6.147}$$

Derive an expression for the second variation of $J[y]$.

**6.6.3. A falling body with drag.** Consider a small, heavy, falling body that experiences air resistance. Assume that the positive direction is downward. Solve the first-order differential equation for this body,

$$m \frac{dv}{dt} \equiv m\,g - b\,v^2\,, \tag{6.148}$$

as a Riccati equation (rather than, for example, as a separable equation). Hint: Use the fact that the terminal velocity is a particular solution.

**6.6.4. Legendre and Jacobi conditions.** Minimize the integral

$$J[y] = \int_1^2 x^2 y'^2 \, dx \tag{6.149}$$

with the boundary conditions

$$y(1) = 0\,, \quad y(2) = 1\,. \tag{6.150}$$

Apply the Legendre condition and the Jacobi condition to determine if your solution is a weak minimum. Determine and draw the one-parameter family of extremals that emanate out of the left boundary point and determine the envelope of this family of extremals. What can you conclude?

**6.6.5. Parabola of safety.** For a simple projectile, fired from the origin, the Lagrangian is

$$L = \frac{1}{2}\,m\,(\dot{x}^2 + \dot{y}^2) - m\,g\,y\,, \tag{6.151}$$

where $g$ is the acceleration due to gravity.

(a) Use Hamilton's principle to derive the equations of motion of the projectile.

(b) Show that the trajectory, corresponding to the initial conditions

$$x(0) = 0\,, \quad y(0) = 0\,, \quad \dot{x}(0) = V\cos\alpha\,, \quad \dot{y}(0) = V\sin\alpha\,, \tag{6.152}$$

is the parabola

$$y(x) = x\,\tan\alpha - \frac{g\,\sec^2\alpha}{2V^2}\,x^2\,. \tag{6.153}$$

(c) Find the envelope of the parabolas that arise as you vary the launch angle $\alpha$ while holding the launch speed $V$ fixed.

**6.6.6. Finding conjugates.** Consider the variational problem

$$J[y] \;=\; \int_0^1 \sqrt{y\,(1+y'^2)}\,dx \tag{6.154}$$

with boundary conditions

$$y(0) \;=\; 2 \;\;\text{and}\;\; y(1) \;=\; 5\,. \tag{6.155}$$

Find the extremals that satisfy the given boundary conditions. Determine whether the extremals satisfy Jacobi's necessary condition.

Chapter 7

# Review and Preview

## 7.1. Introduction

We have looked at a number of classical tests that help us find and characterize weak relative minima. Most of these tests are necessary conditions. These same tests also appear, either in their original form or as strengthened conditions, in discussing sufficient conditions for weak relative minima. It can all be very confusing. Let us take this opportunity to examine a series of necessary and sufficient conditions for weak relative minima, under the simplest possible assumptions.

These necessary and sufficient conditions generalize, in natural ways, to problems with more than one dependent variable, more than one independent variable, higher derivatives, constraints, and a variety of other, more complicated, problems. We will briefly examine the conditions for problems with two dependent variables. See Fox (1950) and Forsyth (1927) for details and for further generalizations.

Finally, we will briefly consider some topics that will appear in the second half of this book.

## 7.2. Necessary conditions

Consider the functional

$$J[y] \;=\; \int_a^b f(x, y, y')\; dx \tag{7.1}$$

with the boundary conditions

$$y(a) \;=\; y_a\,, \quad y(b) \;=\; y_b\,. \tag{7.2}$$

In order for a continuously differentiable function $\hat{y}(x)$ to be a minimum relative to all allowable weak variations, we need to satisfy three necessary conditions. These are as follows:

(a) The function $\hat{y}(x)$ must be an extremal. It must satisfy the Euler–Lagrange equation

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) \;=\; 0\,. \tag{7.3}$$

(b) The Legendre condition,

$$\frac{\partial^2 f}{\partial {y'}^2} \;\geq\; 0\,, \tag{7.4}$$

must be satisfied along $\hat{y}(x)$.

(c) The Jacobi condition,

$$\Delta(x, a) \;=\; u_2(a)\, u_1(x) - u_1(a)\, u_2(x) \;\neq\; 0, \tag{7.5}$$

must be satisfied for all values $a < x < b$. Here $u_1(x)$ and $u_2(x)$ are two, linearly independent, solutions of the Jacobi equation,

$$\frac{d}{dx}(R\, u') + (Q' - P)\, u \;=\; 0\,, \tag{7.6}$$

with

$$P \;\equiv\; f_{yy}(x, \hat{y}, \hat{y}')\,, \quad Q \;\equiv\; f_{yy'}(x, \hat{y}, \hat{y}')\,, \quad R \;=\; f_{y'y'}(x, \hat{y}, \hat{y}')\,. \tag{7.7}$$

Equivalently, $\hat{y}(x)$ should not have any conjugate points between $a$ and $b$.

## 7.3. Sufficient conditions

In order for a continuously differentiable function $\hat{y}(x)$ to be a minimum relative to all allowable weak variations, it is sufficient that we satisfy three conditions. These are as follows:

(a) The function $\hat{y}(x)$ is an extremal. It satisfies the Euler–Lagrange equation

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0\,. \tag{7.8}$$

(b) The strengthened Legendre condition,

$$\frac{\partial^2 f}{\partial y'^2} > 0\,, \tag{7.9}$$

is satisfied along $\hat{y}(x)$.

(c) The strengthened Jacobi condition,

$$\Delta(x,a) = u_2(a)\,u_1(x) - u_1(a)\,u_2(x) \neq 0, \tag{7.10}$$

for all $a < x \le b$, is satisfied.

## 7.4. Two dependent variables

For a functional of the form

$$J[y,z] = \int_a^b f(x,y,z,y',z')\;dx \tag{7.11}$$

with two dependent variables, $y(x)$ and $z(x)$, and fixed boundary conditions, most of the concepts we have already introduced generalize in natural ways:

(a) Extremals must now satisfy a coupled system of Euler–Lagrange equations of the form

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0\,, \quad \frac{\partial f}{\partial z} - \frac{d}{dx}\left(\frac{\partial f}{\partial z'}\right) = 0\,. \tag{7.12}$$

(b) For a minimum, our old strengthened Legendre condition is replaced by a new pair of conditions that state that

$$\frac{\partial^2 f}{\partial y'^2} > 0\,, \quad \frac{\partial^2 f}{\partial z'^2} > 0 \tag{7.13}$$

and that

$$\begin{vmatrix} \dfrac{\partial^2 f}{\partial y'^2} & \dfrac{\partial^2 f}{\partial z' \partial y'} \\ \dfrac{\partial^2 f}{\partial y' \partial z'} & \dfrac{\partial^2 f}{\partial z'^2} \end{vmatrix} > 0 \tag{7.14}$$

or, equivalently, that

$$\frac{\partial^2 f}{\partial y'^2}\,\frac{\partial^2 f}{\partial z'^2} - \left(\frac{\partial^2 f}{\partial y'\,\partial z'}\right)^2 > 0 \tag{7.15}$$

along the extremals $\hat{y}(x)$ and $\hat{z}(x)$.

(c) There are two accessory or secondary characteristic equations,

$$\frac{\partial \Omega}{\partial \xi} - \frac{d}{dx}\left(\frac{\partial \Omega}{\partial \dot{\xi}}\right) = 0\,, \quad \frac{\partial \Omega}{\partial \eta} - \frac{d}{dx}\left(\frac{\partial \Omega}{\partial \dot{\eta}}\right) = 0\,, \tag{7.16}$$

that replace the Jacobi equation. Here, $\Omega$ is an ugly but homogeneous quadratic function of the variations $\xi(x)$ and $\eta(x)$ and their derivatives $\xi'(x)$ and $\eta'(x)$. In particular,

$$\begin{aligned} 2\Omega \;=\; & \frac{\partial^2 f}{\partial y^2}\,\xi^2 + 2\,\frac{\partial^2 f}{\partial y\,\partial z}\,\xi\,\eta + \frac{\partial^2 f}{\partial z^2}\,\eta^2 \\ & + 2\left(\frac{\partial^2 f}{\partial y\,\partial y'}\,\xi\,\xi' + \frac{\partial^2 f}{\partial y\,\partial z'}\,\xi\,\eta' + \frac{\partial^2 f}{\partial y'\,\partial z}\,\xi'\,\eta + \frac{\partial^2 f}{\partial z\,\partial z'}\,\eta\,\eta'\right) \\ & + \frac{\partial^2 f}{\partial y'^2}\,\xi'^2 + 2\,\frac{\partial^2 f}{\partial y'\,\partial z'}\,\xi'\,\eta' + \frac{\partial^2 f}{\partial z'^2}\,\eta'^2\,, \end{aligned} \tag{7.17}$$

where the partial derivatives are evaluated along extremals. See Forsyth (1927) for details.

Fortunately, the solutions to these two equations can be derived from the general solutions of the two Euler–Lagrange equations. If the two Euler–Lagrange equations have solutions of the form

$$y(x) = \phi(x, \alpha, \beta, \gamma, \delta)\,, \quad z(x) = \psi(x, \alpha, \beta, \gamma, \delta)\,, \tag{7.18}$$

where $\alpha$, $\beta$, $\gamma$, and $\delta$ are constants of integration, the solutions to the two accessory equations can be written in terms of partial derivatives of the general solutions to the Euler–Lagrange equations,

$$\xi(x) = c_1\phi_\alpha(x) + c_2\phi_\beta(x) + c_3\phi_\gamma(x) + c_4\phi_\delta(x)\,, \tag{7.19}$$

$$\eta(x) = c_1\psi_\alpha(x) + c_2\psi_\beta(x) + c_3\psi_\gamma(x) + c_4\psi_\delta(x)\,, \tag{7.20}$$

where we now, for conciseness, list the independent variable $x$ as the sole argument.

The range of integration, beginning at $a$, must not extend as far as the conjugate of $a$. This conjugate is the smallest value of $\overline{a}$, greater than $a$, that is a root of the equation

$$\Delta(\overline{a},\, a) = \begin{vmatrix} \phi_\alpha(a) & \phi_\beta(a) & \phi_\gamma(a) & \phi_\delta(a) \\ \psi_\alpha(a) & \psi_\beta(a) & \psi_\gamma(a) & \psi_\delta(a) \\ \phi_\alpha(\bar{a}) & \phi_\beta(\bar{a}) & \phi_\gamma(\bar{a}) & \phi_\delta(\bar{a}) \\ \psi_\alpha(\bar{a}) & \psi_\beta(\bar{a}) & \psi_\gamma(\bar{a}) & \psi_\delta(\bar{a}) \end{vmatrix} = 0\,, \tag{7.21}$$

where the subscripts on $\phi$ and $\psi$ imply partial differentiation with respect to the designated parameter.

In terms of geometry, you should consider a two-parameter family of orbits emanating out of a point. You will, once again, need to look at intersections between nearby orbits.

## 7.5. History and preview

Jacobi's 1837 paper was, in many ways, the high point of the classical phase of the calculus of variations. During this phase, mathematicians implicitly, and perhaps unconsciously, assumed that they were dealing with (once or twice) continuously differentiable extremals and weak variations. During the second half of the 19th century, mathematicians such as Isaac Todhunter, G. Erdmann, and Karl Weierstrass reexamined the assumptions underlying the calculus of variations, put the calculus of variations on a more rigorous footing, and learned to deal with corners and strong variations. The role of Weierstrass was particularly important.

Isaac Todhunter (1871), an English mathematician, first drew attention to solutions of variational problems with corners in his 1871 book. At corners, the derivatives of the solutions change discontinuously; Todhunter called the solutions themselves "discontinuous." Unfortunately, Todhunter was unable to derive analytic conditions that held at the corners. Weierstrass found corner conditions for extremals as early as 1865 but did not publish these conditions. (Weierstrass's derivations appeared much later, after his death, when his lectures were published. Nevertheless, Weierstrass's lectures were extremely influential at the time they were delivered.) Corner conditions first appeared in print in a paper by Erdmann (1877). We now talk of the Weierstrass–Erdmann corner conditions. Once we allow solutions to have corners, it is only a small (but important) step to allow nearby curves to have corners or large derivatives. Weierstrass developed this theory of strong variations.

Corner conditions arose historically out of variable-endpoint conditions. As a result, it will take us a little while to get to corner conditions. We will first follow Weierstrass by taking a *parametric* approach. In order to gain greater generality, Weierstrass regarded the variables $x$ and $y$ for a plane curve as functions of a parameter $t$. This is in contrast to the ordinary approach in which $x$ is the independent variable and $y$ is the dependent variable. The approach will seem a little peculiar at first, but it will allow us to derive a particularly nice pair of variable-endpoint conditions. We need variable-endpoint conditions for problems in which boundary conditions are not fixed. With variable-endpoint conditions in hand, we can quickly derive corner conditions. We can then move on to a new necessary condition, due to Weierstrass, for an extremum relative to strong variations. Finally, we can consider field theory and sufficient conditions for an extremum relative to strong variations.

## 7.6. Recommended reading

Kolmogorov and Yushkevich (1998) and Fraser (2003) discuss the contributions of Todhunter, Erdmann, and Weierstrass to the calculus of variations. See Johnson (1996) for a biography of Isaac Todhunter.

Chapter 8

# The Homogeneous Problem

## 8.1. Integrals in parametric form

We will now follow Weierstrass by replacing the explicit representation

$$y = y(x) \tag{8.1}$$

with a parametric representation,

$$x = x(t)\,, \quad y = y(t)\,, \tag{8.2}$$

when minimizing or maximizing a functional. This important generalization enables us to search for our solutions amongst *regular curves,* such as circles and vertical lines, rather than just amongst *functions.*

Let us consider our simplest integral,

$$J[y] = \int_a^b f(x, y, y')\, dx\,, \tag{8.3}$$

with the boundary conditions

$$y(a) = y_a\,, \quad y(b) = y_b\,. \tag{8.4}$$

We will now let

$$x = x(t)\,, \quad y = y(t) \tag{8.5}$$

for $t_a \le t \le t_b$ with

$$x(t_a) = a\,,\;\; y(t_a) = y_a\,,\;\; x(t_b) = b\,,\;\; y(t_b) = y_b\,. \tag{8.6}$$

For future convenience, we will assume that $x(t)$ and $y(t)$ are continuously differentiable functions whose derivatives, $\dot{x}(t)$ and $\dot{y}(t)$, do not simultaneously vanish,

$$\dot{x}^2 + \dot{y}^2 \; > \; 0\,, \tag{8.7}$$

anywhere on our closed interval $t_a \le t \le t_b$.

Since

$$y' = \frac{dy}{dx} = \frac{\dot{y}}{\dot{x}}\,, \quad dx = \dot{x}\,dt\,, \tag{8.8}$$

integral (8.3) can now be replaced by

$$J[\gamma] = \int_{t_a}^{t_b} f\left(x(t), y(t), \frac{\dot{y}(t)}{\dot{x}(t)}\right) \dot{x}(t)\; dt\,, \tag{8.9}$$

where $\gamma(t) = [x(t), y(t)]$ is the curve corresponding to our parameterization. We are now dealing with a problem with two dependent variables. Our new integrand,

$$F(x, y, \dot{x}, \dot{y}) \;\equiv\; f\left(x(t), y(t), \frac{\dot{y}(t)}{\dot{x}(t)}\right) \dot{x}(t)\,, \tag{8.10}$$

does, moreover, have two special properties:

(a) $F$ does not have any explicit dependence on the independent variable $t$.

(b) $F$ is *homogeneous of degree one* in the two derivatives $\dot{x}$ and $\dot{y}$ in the sense that

$$\begin{aligned} F(x, y, k\dot{x}, k\dot{y}) &= f\left(x(t), y(t), \frac{k\dot{y}(t)}{k\dot{x}(t)}\right) k\dot{x}(t) \\ &= k\,f\left(x(t), y(t), \frac{\dot{y}(t)}{\dot{x}(t)}\right) \dot{x}(t) \\ &= k\,F(x, y, \dot{x}, \dot{y}) \end{aligned} \tag{8.11}$$

for all $k$.

In addition to direct parametric analogs of our simplest nonparametric problem, we will, on occasion, want to consider problems that are formulated, from the start, as parametric problems. In some of these problems, the integrand is *not* homogeneous of degree one but is *positively homogeneous of degree one* in its two derivatives. For example, in many geometric problems the arc length may be written

$$\int_{t_a}^{t_b} F(x, y, \dot{x}, \dot{y})\ dt \ = \ \int_{t_a}^{t_b} \sqrt{\dot{x}^2 + \dot{y}^2}\ dt\,. \tag{8.12}$$

In this case, the integrand satisfies

$$F(x, y, k\dot{x}, k\dot{y}) \ = \ |k|\,\sqrt{\dot{x}^2 + \dot{y}^2}\,. \tag{8.13}$$

Thus,

$$F(x, y, k\dot{x}, k\dot{y}) \ = \ k\,\sqrt{\dot{x}^2 + \dot{y}^2} \tag{8.14}$$

for $k > 0$, but

$$F(x, y, k\dot{x}, k\dot{y}) \ = \ -k\,\sqrt{\dot{x}^2 + \dot{y}^2} \tag{8.15}$$

for $k < 0$. Integrands that satisfy equation (8.14) for $k > 0$ are positively homogeneous of degree one in their derivatives. Integrands that are homogeneous of degree one in their derivatives are automatically positively homogeneous of degree one in their derivatives. In this section, we are only interested in parametric problems that are (at least) positively homogeneous of degree one in their derivatives.

The above properties are important. Let

$$\int_{t_a}^{t_b} F(x, y, \dot{x}, \dot{y})\ dt \tag{8.16}$$

be a functional (a) whose integrand does depend explicitly on $t$ and (b) that is positively homogeneous of degree one in its two derivatives. These problems are called *homogeneous problems,* for obvious reasons. Ordinarily, changing independent variables would have a big effect on the form of an integrand. For our problem, however, let

$$t \ = \ t(\tau) \tag{8.17}$$

be an (orientation-preserving) change of independent variables that satisfies

$$\frac{dt}{d\tau} \ > \ 0 \tag{8.18}$$

for all $t_a < t < t_b$. Clearly,

$$\int_{t_a}^{t_b} F(x, y, \dot{x}, \dot{y})\ dt \ = \ \int_{\tau_a}^{\tau_b} F(x, y, \dot{x}, \dot{y})\, \frac{dt}{d\tau}\, d\tau\,, \tag{8.19}$$

where $t(\tau_a) = t_a$ and $t(\tau_b) = t_b$, and it now follows, using positive homogeneity, that

$$\begin{aligned} \int_{t_a}^{t_b} F(x, y, \dot{x}, \dot{y})\ dt \ &= \ \int_{\tau_a}^{\tau_b} F\left(x, y, \dot{x}\frac{dt}{d\tau}, \dot{y}\frac{dt}{d\tau}\right)\, d\tau \\ &= \ \int_{\tau_a}^{\tau_b} F\left(x, y, \frac{dx}{d\tau}, \frac{dy}{d\tau}\right)\, d\tau\,. \end{aligned} \tag{8.20}$$

Thus, reparameterizing did not change the form of our integrand. As a result, our functional depends only upon the trace of a curve and not on its explicit parametrization. This is good. There may be many equivalent parameterizations for the same simple variational problem and we do not want a simple reparameterization to change our answers.

## 8.2. Euler–Lagrange equations

In one sense, homogeneous problems are easy to solve. Any continuously differentiable solution must satisfy the two Euler–Lagrange equations

$$\frac{\partial F}{\partial x} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}}\right) \ = \ 0\,, \quad \frac{\partial F}{\partial y} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{y}}\right) \ = \ 0\,. \tag{8.21}$$

These two equations cannot, however, be entirely independent. After all, the original nonparametric problem had only a single Euler–Lagrange equation. We will see that the above two Euler–Lagrange equations depend on each other quite strongly — because of the special properties on $F(x, y, \dot{x}, \dot{y})$.

Now, we have seen that $F$ does not have any explicit dependence on $t$. Why then have we not used the traditional first integral

$$\dot{x}\,\frac{\partial F}{\partial \dot{x}} + \dot{y}\,\frac{\partial F}{\partial \dot{y}} - F \ = \ c\,? \tag{8.22}$$

To answer this question, let us begin with a minor detour.

***Euler's identity:***

If $F(x, y, \dot{x}, \dot{y})$ is positively homogeneous of degree one in its derivatives, then $F$ satisfies

$$\dot{x}\, F_{\dot{x}}(x, y, \dot{x}, \dot{y}) + \dot{y}\, F_{\dot{y}}(x, y, \dot{x}, \dot{y}) \;=\; F(x, y, \dot{x}, \dot{y})\,. \tag{8.23}$$

**Proof.** If we start with our homogeneity condition,

$$F(x, y,\, k\dot{x},\, k\dot{y}) \;=\; k\, F(x,\, y,\, \dot{x},\, \dot{y})\,, \tag{8.24}$$

and differentiate with respect to $k$, we obtain

$$\dot{x}\, F_{\dot{x}}(x, y, k\dot{x}, k\dot{y}) + \dot{y}\, F_{\dot{y}}(x, y, k\dot{x}, k\dot{y}) \;=\; F(x, y, \dot{x}, \dot{y})\,. \tag{8.25}$$

This equation is true for all $k > 0$, and so it is true for any $k > 0$. If we now set $k = 1$, we have our identity. ♣

If the independent variable $t$ does not appear explicitly in our integrand, the expression

$$\dot{x}\, F_{\dot{x}} + \dot{y}\, F_{\dot{y}} - F \tag{8.26}$$

is a constant of motion. For a homogeneous problem, however, this expression is, by Euler's identity, *identically* zero. The statement that zero is a constant is true, but not particularly informative.

Although expression (8.26) did not buy us a meaningful first integral, it is still helpful. Using straightforward differentiation (see Exercise 8.6.1), one can easily show that

$$\frac{d}{dt}\left(\dot{x}\, F_{\dot{x}} + \dot{y}\, F_{\dot{y}} - F\right) \;=\; \dot{x}\left(\frac{d}{dt}F_{\dot{x}} - F_x\right) + \dot{y}\left(\frac{d}{dt}F_{\dot{y}} - F_y\right). \tag{8.27}$$

For the homogeneous problem, the left-hand side is identically zero so that

$$\dot{x}\left(\frac{d}{dt}F_{\dot{x}} - F_x\right) + \dot{y}\left(\frac{d}{dt}F_{\dot{y}} - F_y\right) \;=\; 0\,. \tag{8.28}$$

This equation is always true, not just along extremals. So, if a pair of functions $x(t)$ and $y(t)$ satisfies the first Euler–Lagrange equation, it will *automatically* satisfy the second Euler–Lagrange equation, unless $\dot{y} = 0$. Likewise, if this pair satisfies the second Euler–Lagrange equation, it will *automatically* satisfy the first Euler–Lagrange equation, unless $\dot{x} = 0$. The two Euler–Lagrange equations do define a curve, but they do not, because of their lack of independence, pick out a particular parametric representation for that curve. (In addition, you should hesitate to accept any horizontal or vertical line as a solution until you have verified that it satisfies *both* Euler–Lagrange equations.)

Although the two Euler–Lagrange equations are dependent, one Euler–Lagrange equation may still be much easier to work with than the other. Later, we will see an example where it is convenient to work with both Euler–Lagrange equations simultaneously. For the moment, however, let us see if we can replace our two dependent Euler–Lagrange equations with a single symmetric equation.

## 8.3. The Weierstrass equation

Weierstrass reasoned that since the two Euler–Lagrange equations are not independent, he could replace them with a single equation that is symmetric in $x$ and $y$. There is nothing pretty about the derivation of this symmetric equation, but it does provide an interesting alternative to the two Euler–Lagrange equations.

Let us start then with Euler's identity,

$$\dot{x}\, F_{\dot{x}} + \dot{y}\, F_{\dot{y}} \;=\; F\,. \tag{8.29}$$

If we take the partial derivative of this equation with respect to $\dot{x}$, it quickly follows that

$$\dot{x}\,\frac{\partial^2 F}{\partial \dot{x}^2} + \dot{y}\,\frac{\partial^2 F}{\partial \dot{x}\,\partial \dot{y}} \;=\; 0\,. \tag{8.30}$$

If we instead take the partial derivative with respect to $\dot{y}$, we get

$$\dot{x}\,\frac{\partial^2 F}{\partial \dot{x}\,\partial \dot{y}} + \dot{y}\,\frac{\partial^2 F}{\partial \dot{y}^2} \;=\; 0\,. \tag{8.31}$$

From the first of these two equations, we have that

$$\frac{\partial^2 F}{\partial \dot{x}^2} : \frac{\partial^2 F}{\partial \dot{x}\,\partial \dot{y}} \;=\; \frac{\dot{y}}{-\dot{x}} \;=\; \frac{\dot{y}^2}{-\dot{x}\dot{y}} \tag{8.32}$$

while, from the second,

$$\frac{\partial^2 F}{\partial \dot{x}\, \partial \dot{y}} : \frac{\partial^2 F}{\partial \dot{y}^2} \;=\; \frac{-\dot{y}}{\dot{x}} \;=\; \frac{-\dot{x}\dot{y}}{\dot{x}^2}\,. \tag{8.33}$$

Thus

$$\frac{\partial^2 F}{\partial \dot{x}^2} : \frac{\partial^2 F}{\partial \dot{x}\, \partial \dot{y}} : \frac{\partial^2 F}{\partial \dot{y}^2} \;=\; \dot{y}^2 : -\dot{x}\dot{y} : \dot{x}^2\,. \tag{8.34}$$

It now follows that there is a function $F_1$ such that

$$\frac{\partial^2 F}{\partial \dot{x}^2} \;=\; F_1 \dot{y}^2\,, \quad \frac{\partial^2 F}{\partial \dot{x}\, \partial \dot{y}} \;=\; -F_1 \dot{x}\dot{y}\,, \quad \frac{\partial^2 F}{\partial \dot{y}^2} \;=\; F_1 \dot{x}^2 \tag{8.35}$$

and that

$$F_1(x, y, \dot{x}, \dot{y}) \;=\; \frac{F_{\dot{x}\dot{x}} + F_{\dot{y}\dot{y}}}{\dot{x}^2 + \dot{y}^2} \;=\; -\frac{F_{\dot{x}\dot{y}}}{\dot{x}\dot{y}} \tag{8.36}$$

(with $\dot{x}^2 + \dot{y}^2 > 0$). The function $F_1$ will play an important role in our final equation.

Let us now take one of our Euler–Lagrange equations, say

$$\frac{\partial F}{\partial x} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}}\right) \;=\; 0\,. \tag{8.37}$$

Since $F$ has both $x$ and $y$ as dependent variables, the ultradifferentiated form of this equation is

$$F_x - F_{\dot{x}x}\, \dot{x} - F_{\dot{x}y}\, \dot{y} - F_{\dot{x}\dot{x}}\, \ddot{x} - F_{\dot{x}\dot{y}}\, \ddot{y} \;=\; 0. \tag{8.38}$$

Since the partial derivative, with respect to $x$, of Euler's identity is

$$F_x \;=\; F_{x\dot{x}}\, \dot{x} + F_{x\dot{y}}\, \dot{y}\,, \tag{8.39}$$

we substitute this into our ultradifferentiated equation to obtain

$$(F_{x\dot{y}} - F_{\dot{x}y})\, \dot{y} - F_{\dot{x}\dot{x}}\, \ddot{x} - F_{\dot{x}\dot{y}}\, \ddot{y} \;=\; 0\,. \tag{8.40}$$

Finally if we substitute for $F_{\dot{x}\dot{x}}$ and $F_{\dot{x}\dot{y}}$ in terms of $F_1$, we obtain

$$\dot{y}\,[F_{x\dot{y}} - F_{\dot{x}y} + F_1(\dot{x}\ddot{y} - \dot{y}\ddot{x})] \;=\; 0\,. \tag{8.41}$$

If we had started with the other Euler–Lagrange equation, we would have concluded that

$$-\dot{x}\,[F_{x\dot{y}} - F_{\dot{x}y} + F_1(\dot{x}\ddot{y} - \dot{y}\ddot{x})] \;=\; 0 \tag{8.42}$$

and since

$$\dot{x}^2 + \dot{y}^2 \;>\; 0\,, \tag{8.43}$$

by assumption, we conclude that

$$F_{x\dot{y}} - F_{\dot{x}y} + F_1(\dot{x}\ddot{y} - \dot{y}\ddot{x}) \;=\; 0\,. \tag{8.44}$$

This is Weierstrass's symmetric form of the Euler–Lagrange equation for the parametric problem.

To use the Weierstrass equation, we can, for example, express $x$ as any convenient function of $t$, substitute $x(t)$ into equation (8.44), and then solve for $y$ in terms of $t$. The geometrical meaning of the equation is that it determines the curvature

$$\kappa \;=\; \frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{(\dot{x}^2 + \dot{y}^2)^{3/2}} \tag{8.45}$$

when the position, $(x, y)$, and the slope, $\dot{y}/\dot{x}$, are known. Indeed, for an extremal,

$$\kappa \;=\; -\frac{F_{x\dot{y}} - F_{y\dot{x}}}{F_1(\dot{x}^2 + \dot{y}^2)^{3/2}}\,. \tag{8.46}$$

We have only talked about the first variation for homogeneous problems. We can also attack the second variation for a parametric formulation. For example, we can show that the usual Legendre condition, that

$$\frac{\partial^2 f}{\partial {y'}^2} \;\geq\; 0 \tag{8.47}$$

for a minimum, is now replaced by the condition that

$$F_1(x, y, \dot{x}, \dot{y}) \;\equiv\; \frac{F_{\dot{x}\dot{x}} + F_{\dot{y}\dot{y}}}{\dot{x}^2 + \dot{y}^2} \;\geq\; 0 \tag{8.48}$$

for a minimum. See Bolza (1973) or Hancock (1904) for more about the second variation for the homogeneous problem.

**Example 8.1** (Geodesics in the plane)**.**

Consider the integral

$$J[\gamma] \;=\; \int_{t_a}^{t_b} \sqrt{\dot{x}^2 + \dot{y}^2}\; dt \tag{8.49}$$

subject to the boundary conditions

$$x(t_a) \;=\; a\,, \quad y(t_a) \;=\; y_a\,, \quad x(t_b) \;=\; b\,, \quad y(t_b) \;=\; y_b\,. \tag{8.50}$$

The integrand $F$ is clearly positively homogeneous of the first degree in the derivatives.

Let us quickly verify an important claim made earlier. Since

$$\dot{x}\,F_{\dot{x}} + \dot{y}\,F_{\dot{y}} - F \;=\; \frac{\dot{x}^2}{\sqrt{\dot{x}^2+\dot{y}^2}} + \frac{\dot{y}^2}{\sqrt{\dot{x}^2+\dot{y}^2}} - \sqrt{\dot{x}^2+\dot{y}^2}\,, \qquad (8.51)$$

it is, in fact, the case that

$$\dot{x}\,F_{\dot{x}} + \dot{y}\,F_{\dot{y}} - F \;=\; 0\,. \qquad (8.52)$$

Since $x$ and $y$ do not appear explicitly in our integrand, the two Euler–Lagrange equations reduce to

$$\frac{\partial F}{\partial \dot{x}} \;=\; \frac{\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}} \;=\; \alpha\,, \quad \frac{\partial F}{\partial \dot{y}} \;=\; \frac{\dot{y}}{\sqrt{\dot{x}^2+\dot{y}^2}} \;=\; \beta\,, \qquad (8.53)$$

where $\alpha$ and $\beta$ are constants. A trivial solution of the first Euler–Lagrange equation is $\dot{y} = 0$. This also satisfies the second Euler–Lagrange equation (although it did not have to). Another trivial solution of the first Euler–Lagrange equation is $\dot{x} = 0$. This trivial solution did have to satisfy the second Euler–Lagrange equation. These trivial solutions give us horizontal and vertical lines.

More generally,

$$\frac{\left(\frac{\dot{y}}{\sqrt{\dot{x}^2+\dot{y}^2}}\right)}{\left(\frac{\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}}\right)} \;=\; \frac{dy}{dx} \;=\; \frac{\beta}{\alpha} \;=\; m\,, \qquad (8.54)$$

which gives us the lines

$$y \;=\; m\,x + c \qquad (8.55)$$

as possible extremals. For a real problem, we would now use our boundary conditions to determine $m$ and $c$.

The Weierstrass equation for this problem is

$$\frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{(\dot{x}^2+\dot{y}^2)^{3/2}} \;=\; 0\,; \qquad (8.56)$$

our extremals must have zero curvature, $\kappa = 0$, everywhere. We conclude, once again, that our extremals are lines. If we arbitrarily set

$$x(t) \;=\; t\,, \qquad (8.57)$$

then the Weierstrass equation reduces to

$$\ddot{y} = 0. \tag{8.58}$$

Two integrations give us

$$y = m\,t + c \tag{8.59}$$

or

$$y = m\,x + c. \tag{8.60}$$

## 8.4. Case study: The parametric Queen Dido problem

Consider the problem of maximizing the circumscribed area of a closed curve,

$$J[x,y] = \int_{t_0}^{t_1} \frac{1}{2}\,(x\,\dot{y} - y\,\dot{x})\;dt\,, \tag{8.61}$$

subject to the boundary conditions

$$x(t_0) = +a,\;\; y(t_0) = 0,\;\; x(t_1) = -a,\;\; y(t_1) = 0 \tag{8.62}$$

and the integral constraint

$$K[x,y] = \int_{t_0}^{t_1} \sqrt{\dot{x}^2 + \dot{y}^2}\;dt = l\,. \tag{8.63}$$

There is only one constraint in this problem and so we expect one Lagrange multiplier. Let us therefore introduce

$$J - \lambda\,K = \int_{t_0}^{t_1} \left[\frac{1}{2}(x\,\dot{y} - y\,\dot{x}) - \lambda\sqrt{\dot{x}^2 + \dot{y}^2}\right]\;dt\,. \tag{8.64}$$

Since this integrand is positively homogeneous of degree one in its derivatives, we are dealing with a homogeneous problem.

The Weierstrass equation, you will remember, takes the form

$$F_{x\dot{y}} - F_{y\dot{x}} + F_1(\dot{x}\ddot{y} - \dot{y}\ddot{x}) = 0\,, \tag{8.65}$$

where

$$F_1 = \frac{F_{\dot{x}\dot{x}} + F_{\dot{y}\dot{y}}}{\dot{x}^2 + \dot{y}^2} = -\frac{F_{\dot{x}\dot{y}}}{\dot{x}\dot{y}}\,. \tag{8.66}$$

For our integrand,

$$F_{x\dot{y}} = \frac{1}{2}, \quad F_{y\dot{x}} = -\frac{1}{2}, \tag{8.67}$$

and

$$F_1 = -\frac{\lambda}{(\dot{x}^2 + \dot{y}^2)^{3/2}}, \tag{8.68}$$

so that

$$1 - \lambda \frac{(\dot{x}\ddot{y} - \dot{y}\ddot{x})}{(\dot{x}^2 + \dot{y}^2)^{3/2}} = 0 \tag{8.69}$$

or

$$\kappa = \frac{1}{\lambda}. \tag{8.70}$$

Since the extremals have a constant curvature of one over $\lambda$, they also have a constant radius of curvature of $\lambda$ and are circles of radius $\lambda$.

We can draw the same conclusions from the two Euler–Lagrange equations

$$\frac{1}{2}\dot{y} + \frac{1}{2}\frac{d}{dt}(y) - \lambda\left[-\frac{d}{dt}\left(\frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}}\right)\right] = 0, \tag{8.71}$$

$$-\frac{1}{2}\dot{x} - \frac{1}{2}\frac{d}{dt}(x) - \lambda\left[-\frac{d}{dt}\left(\frac{\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}}\right)\right] = 0. \tag{8.72}$$

If we multiply through by $dt/ds$, where $s$ is arc length, the two Euler–Lagrange equations quickly simplify to

$$\frac{dy}{ds} + \lambda\frac{d^2x}{ds^2} = 0, \quad -\frac{dx}{ds} + \lambda\frac{d^2y}{ds^2} = 0. \tag{8.73}$$

These equations have the solution

$$x = \alpha + \gamma \sin\left(\frac{s}{\lambda} + \delta\right), \quad y = \beta - \gamma \cos\left(\frac{s}{\lambda} + \delta\right), \tag{8.74}$$

where $\alpha$, $\beta$, $\gamma$, and $\delta$ are four constants of integration. This solution corresponds to a circle of radius $\gamma$ centered at $(\alpha, \beta)$. For fixed $a$, different perimeters $l$ determine different circular arcs.

From the symmetry of the problem, it is clear that $\alpha = 0$. Applying the constraint to the solution makes it clear that $\lambda = \gamma$, as expected. The solutions now take the simpler form

$$x = \gamma \sin\left(\frac{s}{\gamma} + \delta\right), \quad y = \beta - \gamma \cos\left(\frac{s}{\gamma} + \delta\right). \tag{8.75}$$

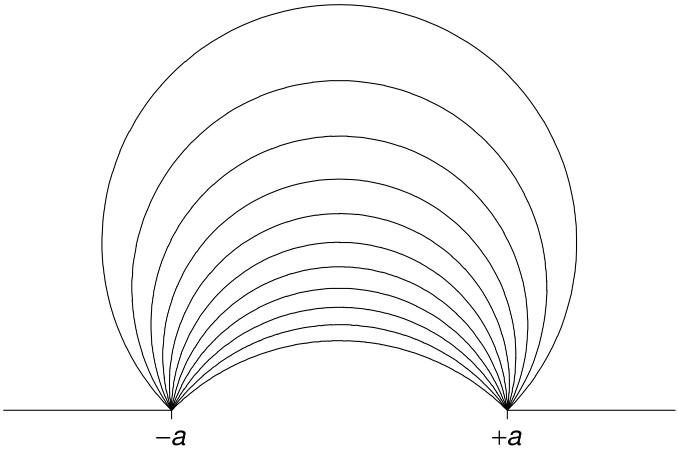



**Figure 8.1.** Different circular arcs

The boundary conditions, in turn, imply that

$$a \;=\; \gamma \,\sin\,\delta\,, \quad -a \;=\; \gamma \,\sin\left(\frac{l}{\gamma}+\delta\right) \tag{8.76}$$

and

$$\beta \;=\; \gamma \,\cos\,\delta\,, \quad \beta \;=\; \gamma \,\cos\left(\frac{l}{\gamma}+\delta\right)\,. \tag{8.77}$$

It now follows that

$$\frac{l}{\gamma}+\delta \;=\; 2\pi-\delta \tag{8.78}$$

so that

$$\delta \;=\; \pi-\frac{l}{2\gamma}\,. \tag{8.79}$$

Our solution may now be written as

$$x \;=\; \gamma \,\sin\left(\frac{l}{2\gamma}-\frac{s}{\gamma}\right)\,, \quad y \;=\; \beta+\gamma \,\cos\left(\frac{l}{2\gamma}-\frac{s}{\gamma}\right) \tag{8.80}$$

(see Figure 8.1) and our boundary conditions are now just

$$a \;=\; \gamma \,\sin\left(\frac{l}{2\gamma}\right)\,, \quad -\beta \;=\; \gamma \,\cos\left(\frac{l}{2\gamma}\right)\,. \tag{8.81}$$

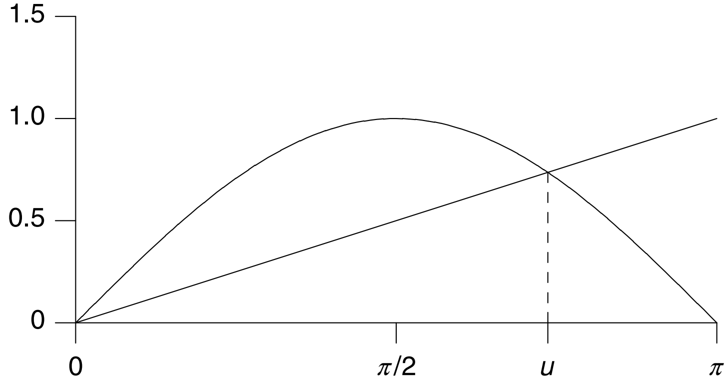



**Figure 8.2.** Graphical solution of $u$

Let

$$u \equiv \frac{l}{2\gamma} \tag{8.82}$$

so that

$$\gamma = \frac{l}{2u}\,. \tag{8.83}$$

The first boundary condition may now be rewritten as

$$\frac{2a}{l}\,u = \sin u\,. \tag{8.84}$$

This equation can be solved graphically (see Figure 8.2) for $u$ and for the radius $\gamma$. It is clear that $0 < u \leq \pi$ if $l > 2a$. With $u$ and the radius $\gamma$ in hand, the $y$ coordinate of the center of the circle is simply

$$\beta = -\frac{l}{2}\,\frac{\cos u}{u}\,. \tag{8.85}$$

## 8.5. Recommended reading

Several books cover the homogeneous or parametric problem of the calculus of variations. See, for example, the books authored by Bolza (1973), Brechtken-Manderscheid (1991), Clegg (1968), Hancock (1904), Pars (1962), and Sagan (1969). The book by Hancock (1904) is based on the lectures of H. A. Schwarz and Karl Weierstrass and takes a strongly parametric approach to the calculus of variations.

Bliss (1916) and Hestenes (1934) analyze the second variation and Jacobi's condition for parametric problems.

In classical mechanics, the Hamiltonian is identically zero for homogeneous problems. As a result, one can no longer derive the canonical equations in the usual way. Forbes (1991) reviews proposed solutions to this problem.

Parametric problems are especially common in optics. See, for example, Stavroudis (2006).

## 8.6. Exercises

**8.6.1. Some differentiation.** Show that

$$\frac{d}{dt}\left(\dot{x}\,F_{\dot{x}} + \dot{y}\,F_{\dot{y}} - F\right) \;=\; \dot{x}\left(\frac{d}{dt}F_{\dot{x}} - F_x\right) + \dot{y}\left(\frac{d}{dt}F_{\dot{y}} - F_y\right). \quad (8.86)$$

**8.6.2. Minimal surface of revolution revisited.** Reformulate the problem of the minimal surface of revolution (the soap-film problem) as a homogeneous problem. Why might one of the two Euler–Lagrange equations for this problem be better than the other? Find the extremal arc for this (homogeneous) problem using your Euler–Lagrange equations. What is the Weierstrass equation for this problem?

**8.6.3. Weierstrass equation.** Determine $F_1(x, y, \dot{x}, \dot{y})$ and find the Weierstrass equation for the following integrands:

(a)

$$F(x, y, \dot{x}, \dot{y}) \;=\; n(x,\, y)\,\sqrt{\dot{x}^2 + \dot{y}^2}\,, \quad (8.87)$$

(b)

$$F(x, y, \dot{x}, \dot{y}) \;=\; \frac{\dot{y}^2}{\dot{x}}\,, \quad (8.88)$$

(c)

$$F(x, y, \dot{x}, \dot{y}) \;=\; \frac{\dot{x}^2 + \dot{y}^2}{\sqrt{2(\dot{x}^2 + \dot{y}^2)} + \dot{x}}\,. \quad (8.89)$$

Chapter 9

# Variable-Endpoint Conditions

## 9.1. Natural boundary conditions

In most of the problems that we have considered, we have specified both boundary conditions. Sometimes, you will need to determine one or more boundary conditions as part of the optimization problem.

**Example 9.1** (Zermelo's navigation problem)**.**

Consider a boat crossing a river (see Figure 9.1). The river has parallel and straight banks, $b$ units apart. We will take the left bank to be the $y$-axis. We will assume that the downstream current has a speed that depends on the $x$ coordinate,

$$v = v(x)\,, \tag{9.1}$$

and that the boat has a constant speed $c$ ($c > v$) in still water so that the rate of change of position of the boat is

$$\frac{dx}{dt} = c\cos\theta\,, \tag{9.2}$$

$$\frac{dy}{dt} = v(x) + c\sin\theta\,, \tag{9.3}$$

where

$$\theta = \theta(x) \tag{9.4}$$

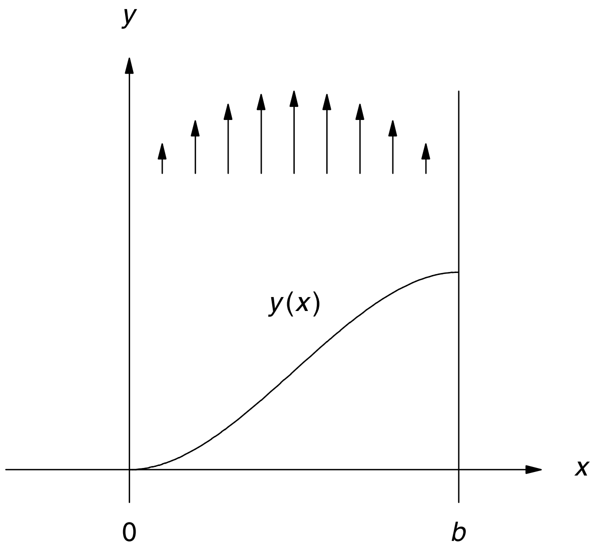



**Figure 9.1.** Crossing a river

is the steering angle of the boat. Let us imagine that the boat starts at the origin. What curve,

$$y = y(x)\,, \tag{9.5}$$

minimizes the travel time across the river?

The crossing time satisfies

$$T = \int_0^b \frac{dt}{dx}\,dx = \int_0^b \frac{1}{c\,\cos\,\theta}\,dx\,. \tag{9.6}$$

To put this problem into standard form, we must rewrite our integrand as a function of $v$, $c$, and, most importantly, $y'$. Since

$$y' = \frac{dy/dt}{dx/dt} = \frac{v + c\,\sin\,\theta}{c\,\cos\,\theta}\,, \tag{9.7}$$

we may write

$$y' \cdot c\,\cos\,\theta - v = \pm\,c\,\sqrt{1 - \cos^2\theta}\,. \tag{9.8}$$

After squaring both sides,

$$c^2 \cos^2\theta\, {y'}^2 - 2\,v\,c\,\cos\theta\, y' + v^2 \;=\; c^2(1-\cos^2\theta)\,, \tag{9.9}$$

and simplifying, we get

$$(c^2 - v^2)\frac{1}{\cos^2\theta} + 2\,v\,y'\,c\,\frac{1}{\cos\,\theta} - c^2\,(1+{y'}^2) \;=\; 0\,. \tag{9.10}$$

Solving for the reciprocal of $\cos\,\theta$, we find that

$$T \;=\; \int_0^b \frac{\sqrt{c^2\,(1+{y'}^2) - v^2(x)} - v(x)\,y'}{c^2 - v^2(x)}\,dx\,, \tag{9.11}$$

where $v(x)$ is a known function of $x$.

We have only specified one boundary condition for this problem, at the beginning point

$$y(0) \;=\; 0\,. \tag{9.12}$$

The endpoint can be anywhere along the opposite bank, $x = b$. Different terminal points will yield different smallest crossing times. Part of the problem is to choose the right terminal point. After all, the problem, as posed, is simply to cross the river as quickly as possible. We're willing to land anywhere to achieve this goal.

With this example in mind, let us minimize

$$J[y] \;=\; \int_a^b f(x,y,y')\;dx \tag{9.13}$$

with the $x$ coordinates of the endpoints given, but the $y$ coordinates unspecified.

We have seen that the first variation is just

$$\delta J \;=\; \epsilon\int_a^b [f_y(x,\hat{y},\hat{y}')\,\eta + f_{y'}\,(x,\hat{y},\hat{y}')\,\eta']\;dx\,. \tag{9.14}$$

We will, as usual, integrate by parts. For convenience, let us follow Lagrange's approach. It now follows that

$$\int_a^b f_{y'}(x,\hat{y},\hat{y}')\,\eta'\;dx \;=\; \eta(x)\,\frac{\partial f}{\partial y'}\bigg|_{x\,=\,a}^{x\,=\,b} - \int_a^b \eta\,\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\,dx \tag{9.15}$$

so that the first variation takes the form

$$\delta J \;=\; \epsilon\,\eta(x)\,\frac{\partial f}{\partial y'}\bigg|_{x\,=\,a}^{x\,=\,b} + \epsilon\int_a^b \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right]\eta(x)\;dx\,. \tag{9.16}$$

The first variation must vanish for *all* $\eta(x)$. Since this includes all $\eta(x)$ that vanish at the endpoints, the Euler–Lagrange equation must still be satisfied. If the left boundary condition is unspecified, $\eta(x)$ need not vanish at $x = a$ and we instead require that

$$\frac{\partial f}{\partial y'}(a, \hat{y}(a), \hat{y}'(a)) \;=\; 0\,. \tag{9.17}$$

Similarly, if the right boundary condition is unspecified, $\eta(x)$ need not vanish at $x = b$ and we demand that

$$\frac{\partial f}{\partial y'}(b, \hat{y}(b), \hat{y}'(b)) \;=\; 0\,. \tag{9.18}$$

These two equations are often referred to as *natural boundary conditions.* This is because these two conditions arise "naturally" in a problem that, at the outset, is not equipped with boundary conditions.

**Example 9.2** (Zermelo's problem (continued))**.**

Let us now find the trajectory $y = y(x)$ that minimizes the travel time,

$$T \;=\; \int_0^1 f(x, y, y')\;dx \tag{9.19}$$

$$=\; \int_0^1 \frac{\sqrt{c^2\,(1 + y'^2) - v^2(x)} - v(x)\,y'}{c^2 - v^2(x)}\;dx\,,$$

across a river of width $b = 1$.

Since the integrand is independent of $y$, the Euler–Lagrange equation reduces to

$$\frac{\partial f}{\partial y'} \;=\; \alpha \tag{9.20}$$

with $\alpha$ a constant. Thus

$$\frac{1}{c^2 - v^2}\left[\frac{c^2\, y'}{\sqrt{c^2\,(1 + y'^2) - v^2(x)}} - v\right] \;=\; \alpha\,. \tag{9.21}$$

However, since the right boundary condition is unspecified, the natural boundary condition implies that

$$\left.\frac{\partial f}{\partial y'}\right|_{x\,=\,1} \;=\; 0 \tag{9.22}$$

so that $\alpha = 0$ and

$$\frac{1}{c^2 - v^2}\left[\frac{c^2\, y'}{\sqrt{c^2\,(1 + y'^2) - v^2(x)}} - v\right] \;=\; 0\,. \tag{9.23}$$

This simplifies to

$$c\,y' \;=\; \pm v(x)\,. \tag{9.24}$$

For the positive root, the crossing time,

$$T \;=\; \frac{1}{c}\,, \tag{9.25}$$

is inversely proportional to the speed of the vessel. For the negative root,

$$T \;=\; \frac{1}{c}\int_0^1 \frac{c^2 + v^2(x)}{c^2 - v^2(x)}\;dx\,, \tag{9.26}$$

with an integral that is greater than one. Since we would like to minimize the crossing time, we will take the positive root.

For a parabolic velocity profile,

$$v(x) \;=\; x\,(1 - x)\,, \tag{9.27}$$

integrating the positive $y'$ root produces

$$\begin{aligned} y \;&=\; \frac{1}{c}\int v(x)\;dx \;=\; \frac{1}{c}\int (x - x^2)\;dx \\ &=\; \frac{1}{c}\left(\frac{x^2}{2} - \frac{x^3}{3}\right) + \beta\,, \end{aligned} \tag{9.28}$$

but, from the boundary condition at $(0, 0)$, $\beta = 0$ and

$$y \;=\; \frac{1}{c}\left(\frac{x^2}{2} - \frac{x^3}{3}\right)\,. \tag{9.29}$$

The ideal landing spot is clearly

$$\left(1\,,\frac{1}{6c}\right)\ . \tag{9.30}$$

**Example 9.3.**

Consider a functional of the form

$$J[y] \;=\; \int_a^b k(x,y)\sqrt{1+y'^2}\;dx \tag{9.31}$$

with the left boundary condition specified,

$$y(a) \;=\; y_a\,, \tag{9.32}$$

but with the right boundary condition, at $x = b$, unspecified. Note that setting

$$k(x,\,y) \;=\; \frac{1}{\sqrt{y}} \tag{9.33}$$

leads to the brachistochrone problem while setting

$$k(x,\,y) \;=\; y \tag{9.34}$$

produces the minimal surface of revolution problem.

For general $k(x, y)$,

$$\left.\frac{\partial f}{\partial y'}\right|_{x\,=\,b} \;=\; \left.\frac{k(x,y)y'}{\sqrt{1+y'^2}}\right|_{x\,=\,b} \;=\; 0 \tag{9.35}$$

implies that

$$y'(b) \;=\; 0\,. \tag{9.36}$$

This means that the extremal must hit the boundary $x = b$ at a right angle. This is enough information to determine $y_b$.

We can easily extend our discussion of natural boundary conditions to homogeneous problems in parametric form. For an integrand of the form

$$F(x,y,\dot{x},\dot{y})\,, \tag{9.37}$$

we require

$$\left.\frac{\partial F}{\partial \dot{y}}\right|_{x\,=\,b} \;=\; 0 \tag{9.38}$$

if the solution has to end on the vertical line

$$x = b\,. \tag{9.39}$$

If the solution has to end on the horizontal line

$$y = c\,, \tag{9.40}$$

we instead require that

$$\left.\frac{\partial F}{\partial \dot{x}}\right|_{y=c} = 0\,. \tag{9.41}$$

We will see how these results arise in the next section.

## 9.2. Transversality conditions

Natural boundary conditions allow us to determine the correct boundary condition for an endpoint constrained to lie on the vertical line

$$x = b\,. \tag{9.42}$$

We would like to do better. Ideally, we want to determine the correct boundary condition for an endpoint constrained to lie along some general curve. To put this in the context of Zermelo's navigation problem, few rivers have straight banks. How do we handle rivers with curved banks?

In determining natural boundary conditions, we allowed the ordinate of the endpoint to vary. We now want both the abscissa and the ordinate to vary. It will be easiest if we can do this in a way that treats $x$ and $y$ as equals. So, rather than writing the functional in the ordinary way, we will revert to writing our functional in the parametric form

$$J[\gamma] = \int_{t_a}^{t_b} F(x, y, \dot{x}, \dot{y})\; dt\,. \tag{9.43}$$

$F(x, y, \dot{x}, \dot{y})$ here is, as usual, positively homogeneous of degree one in the two derivatives $\dot{x}$ and $\dot{y}$.

Suppose that our endpoints are constrained to lie on two curves, $C_a$ and $C_b$ (see Figure 9.2). We assume that there is a curve

$$x(t) = \hat{x}(t)\,, \quad y(t) = \hat{y}(t)\,, \quad t_a \le t \le t_b, \tag{9.44}$$

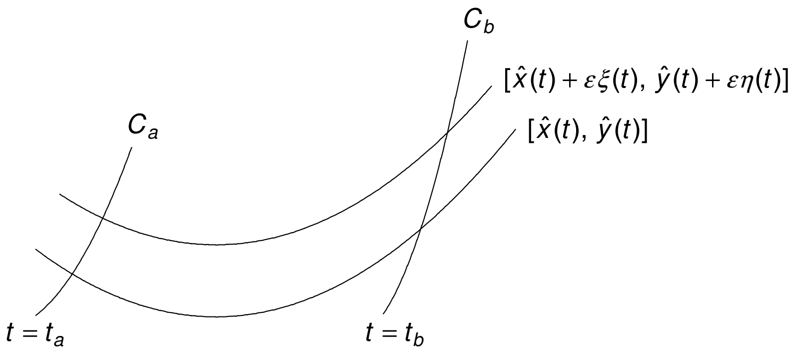



**Figure 9.2.** Variable endpoints

that minimizes our functional and we will now consider small deviations or variations that result in curves,

$$x(t) = \hat{x}(t) + \epsilon\,\xi(t)\,, \tag{9.45}$$

$$y(t) = \hat{y}(t) + \epsilon\,\eta(t)\,, \tag{9.46}$$

that also have their endpoints on $C_a$ and $C_b$.

We will assume that the parametric representation of each varied curve is such that parameter values at the endpoints are always $t = t_a$ and $t = t_b$. This need not be so. A neighboring curve could be defined in terms of the parameter $\tau$ on the interval

$$\tau_a \le \tau \le \tau_b\,. \tag{9.47}$$

Nevertheless, we saw, early in our discussion of the homogeneous problem, that the value of the functional depends only upon the trace of the curve and not on the explicit parameterization. Hence, we can always impose the parameter transformation

$$t = t_a + \frac{(t_b - t_a)\,(\tau - \tau_a)}{\tau_b - \tau_a} \tag{9.48}$$

and force our assumption to be true. The point of this assumption is that we want to consider variations in $x$ and $y$, but not in the independent variable $t$.

Proceeding in the usual manner, we now consider the total variation

$$\Delta J \;=\; \int\limits_{t_a}^{t_b} F(\hat{x}+\epsilon\xi, \hat{y}+\epsilon\eta, \dot{\hat{x}}+\epsilon\dot{\xi}, \dot{\hat{y}}+\epsilon\dot{\eta})\; dt \qquad (9.49)$$

$$-\int\limits_{t_a}^{t_b} F(\hat{x},\, \hat{y},\, \dot{\hat{x}},\, \dot{\hat{y}})\; dt\,.$$

If we expand the total variation in a Taylor series in $\epsilon$ and keep the terms that are first order in $\epsilon$, we obtain the first variation

$$\delta J \;=\; \epsilon \int\limits_{t_a}^{t_b} \left( \frac{\partial F}{\partial x}\,\xi + \frac{\partial F}{\partial \dot{x}}\,\dot{\xi} + \frac{\partial F}{\partial y}\,\eta + \frac{\partial F}{\partial \dot{y}}\,\dot{\eta} \right)\; dt\,. \qquad (9.50)$$

We may integrate two of the terms in this integral by parts following Lagrange's approach,

$$\int\limits_{t_a}^{t_b} \frac{\partial F}{\partial \dot{x}}\,\dot{\xi}\; dt \;=\; \frac{\partial F}{\partial \dot{x}}\,\xi\,\Bigg|_{t\,=\,t_a}^{t\,=\,t_b} - \int\limits_{t_a}^{t_b} \xi\,\frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}}\right)\; dt\,, \qquad (9.51)$$

$$\int\limits_{t_a}^{t_b} \frac{\partial F}{\partial \dot{y}}\,\dot{\eta}\; dt \;=\; \frac{\partial F}{\partial \dot{y}}\,\eta\,\Bigg|_{t\,=\,t_a}^{t\,=\,t_b} - \int\limits_{t_a}^{t_b} \eta\,\frac{d}{dt}\left(\frac{\partial F}{\partial \dot{y}}\right)\; dt\,, \qquad (9.52)$$

to obtain

$$\delta J = \epsilon \int\limits_{t_a}^{t_b} \left\{ \xi \left[ \frac{\partial F}{\partial x} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}}\right) \right] + \eta \left[ \frac{\partial F}{\partial y} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{y}}\right) \right] \right\}\; dt \qquad (9.53)$$

$$+\,\epsilon \left( \frac{\partial F}{\partial \dot{x}}\,\xi + \frac{\partial F}{\partial \dot{y}}\,\eta \right)_{t\,=\,t_a}^{t\,=\,t_b}\,.$$

The first variation must vanish for *all* weak variations, $\xi(t)$ and $\eta(t)$, that lie on the curves $C_a$ and $C_b$. This includes variations that vanish at the endpoints,

$$\xi(t_a) \;=\; \xi(t_b) \;=\; 0\,, \quad \eta(t_a) \;=\; \eta(t_b) \;=\; 0\,, \qquad (9.54)$$

and so we quickly recover the two Euler–Lagrange equations

$$\frac{\partial F}{\partial x} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}}\right) = 0\,, \quad \frac{\partial F}{\partial y} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{y}}\right) = 0\,. \tag{9.55}$$

Thus, for extremals,

$$\delta J = \epsilon \left(\frac{\partial F}{\partial \dot{x}}\,\xi + \frac{\partial F}{\partial \dot{y}}\,\eta\right)_{t=t_a}^{t=t_b}\,. \tag{9.56}$$

Now, if the right endpoint is, for example, constrained to lie on a vertical line, then $\xi(t_b) = 0$, $\eta(t_b)$ is free, and we must impose the condition

$$\left.\frac{\partial F}{\partial \dot{y}}\right|_{t=t_b} = 0\,. \tag{9.57}$$

If, however, the right endpoint is constrained to lie on a horizontal line, then $\eta(t_b) = 0$, $\xi(t_b)$ is free, and we must instead impose the condition

$$\left.\frac{\partial F}{\partial \dot{x}}\right|_{t=t_b} = 0\,. \tag{9.58}$$

In these cases, we obtain the natural boundary conditions.

More generally, the right endpoint may be constrained to lie along the simple curve $C_b$. The quantities $\xi(t_b)$ and $\eta(t_b)$ may both be nonzero, but they are not independent since they correspond to simultaneous changes in $x$ and $y$ along the curve $C_b$. Indeed, if the curve $C_b$ has the parametric representation

$$x = \phi(u)\,, \quad y = \psi(u) \tag{9.59}$$

and if $u_0$ is the value of $u$ such that

$$\phi(u_0) = \hat{x}(t_b)\,, \quad \psi(u_0) = \hat{y}(t_b)\,, \tag{9.60}$$

then

$$\epsilon\,\xi(t_b) \approx \phi'(u_0)\,du\,, \quad \epsilon\,\eta(t_b) \approx \psi'(u_0)\,du \tag{9.61}$$

and

$$\left[\frac{\partial F}{\partial \dot{x}}\,\phi'(u_0) + \frac{\partial F}{\partial \dot{y}}\,\psi'(u_0)\right]_{t=t_b} = 0\,. \tag{9.62}$$

This is a *transversality condition* that may be used to determine the terminal endpoint. An analogous condition may be obtained for a variable beginning point.

A transversality condition for a homogeneous problem can easily be used to derive a transversality condition for a problem in ordinary form. (This is arguably the simplest way of doing so.) Indeed, for $\dot{x} > 0$,

$$F(x, y, \dot{x}, \dot{y}) \;=\; \dot{x}\, F\left(x, y, 1, \frac{\dot{y}}{\dot{x}}\right) \;=\; f(x, y, y')\, \dot{x} \tag{9.63}$$

with the first equality following from the positive homogeneity of $F$ for $\dot{x}$ and $\dot{y}$. (This equality is more generally true for $\dot{x} \neq 0$ for functions $F$ that are homogeneous in $\dot{x}$ and $\dot{y}$.) In the second equality, the third argument of $F$ is, in effect, dropped; the partial of $F$ with respect to its fourth argument is equivalent to the partial of $f$ with respect to its third argument. Thus

$$\begin{aligned} \frac{\partial F}{\partial \dot{x}} &= F\left(x, y, 1, \frac{\dot{y}}{\dot{x}}\right) - \frac{\dot{x}\dot{y}}{\dot{x}^2}\, F_{\dot{y}}\left(x, y, 1, \frac{\dot{y}}{\dot{x}}\right) \\ &= f(x,\, y,\, y') - y'\, \frac{\partial f}{\partial y'}(x, y, y') \end{aligned} \tag{9.64}$$

and

$$\begin{aligned} \frac{\partial F}{\partial \dot{y}} &= \frac{\dot{x}}{\dot{x}}\, F_{\dot{y}}\left(x, y, 1, \frac{\dot{y}}{\dot{x}}\right) \\ &= \frac{\partial f}{\partial y'}(x, y, y')\,. \end{aligned} \tag{9.65}$$

The transversality condition

$$\left[\frac{\partial F}{\partial \dot{x}}\, \phi'(u_0) + \frac{\partial F}{\partial \dot{y}}\, \psi'(u_0)\right]_{t\,=\,t_b} \;=\; 0 \tag{9.66}$$

now takes the form

$$\left[\left(f - y'\, \frac{\partial f}{\partial y'}\right) \phi'(u_0) + \frac{\partial f}{\partial y'}\, \psi'(u_0)\right]_{t\,=\,t_b} \;=\; 0. \tag{9.67}$$

If the terminal curve is specified as a function,

$$y \;=\; g(x)\,, \tag{9.68}$$

rather than parametrically, then

$$\frac{\psi'(u_0)}{\phi'(u_0)} \;=\; g'(b) \;\approx\; \left.\frac{\delta y}{\delta x}\right|_{x\,=\,b} \tag{9.69}$$

for $\hat{x}(t_b) = b$. (We are using $\delta x$ and $\delta y$ to represent the variation in $x$ and $y$ along the boundary curve.) As a result, the transversality condition simplifies to

$$\phi'(u_0) \left[ \left( f - y' \frac{\partial f}{\partial y'} \right) + g' \frac{\partial f}{\partial y'} \right]_{x=b} = 0 \qquad (9.70)$$

or

$$\left[ \left( f - y' \frac{\partial f}{\partial y'} \right) \delta x + \frac{\partial f}{\partial y'} \delta y \right]_{x=b} = 0\,. \qquad (9.71)$$

This transversality condition may be used to determine the terminal endpoint. A similar condition holds for the initial endpoint and the most general version of the transversality condition that we can write down for the standard or ordinary problem is

$$\left[ \left( f - y' \frac{\partial f}{\partial y'} \right) \delta x + \frac{\partial f}{\partial y'} \delta y \right]_{x=a}^{x=b} = 0\,. \qquad (9.72)$$

We can make one final simplification, for problems in classical mechanics. Remember that in a one degree of freedom problem in Lagrangian mechanics, time $t$ is the independent variable, the generalized coordinate $q(t)$ might be our dependent variable, and the Lagrangian $L(t, q, \dot{q})$ is our integrand,

$$x \to t\,, \quad y(x) \to q(t)\,, \quad f(x, y, y') \to L(t, q, \dot{q})\,. \qquad (9.73)$$

Let us now shift to a Hamiltonian formulation, with

$$p = \frac{\partial}{\partial \dot{q}} L(t, q, \dot{q})\,, \qquad (9.74)$$

as the momentum and

$$H(t, q, p) = p\, \dot{q}(t, q, p) - L(t, q, \dot{q}(t, q, p)) \qquad (9.75)$$

as the Hamiltonian. The transversality condition now takes the form

$$(p\, \delta q - H\, \delta t)_{t=t_a}^{t=t_b} = 0\,. \qquad (9.76)$$

This readily generalizes to the case of $n$ dependent variables, $q_i(t)$, with $t$ as the independent variable:

$$\left[ \left( \sum_{i=1}^{n} p_i\, \delta q_i \right) - H\, \delta t \right]_{t=t_a}^{t=t_b} = 0\,. \qquad (9.77)$$

The homogeneous problem can be thought of as a special case of this problem where the Hamiltonian is identically equal to zero. For this special case, we come back, full circle, to a transversality condition that looks like the sum of a set of natural boundary conditions.

We've been looking at quite a bit of theory. So, let us consider a pair of examples. The first example is fairly concrete. The second example is extremely concrete.

**Example 9.4** (Fermat-type integrals)**.**

Consider

$$J[y] \;=\; \int_a^b k(x,y)\sqrt{1+y'^2}\;dx\,. \tag{9.78}$$

You will remember that setting

$$k(x,y) \;=\; \frac{1}{\sqrt{y}} \tag{9.79}$$

leads to the brachistochrone problem, while setting

$$k(x,\,y) \;=\; y \tag{9.80}$$

produces the minimal surface of revolution problem.

For this problem,

$$\frac{\partial f}{\partial y'} \;=\; k(x,y)\frac{y'}{\sqrt{1+y'^2}} \;=\; \frac{y'f}{1+y'^2}\,. \tag{9.81}$$

Thus, the transversality condition at $b$, for a terminal curve $y = g(x)$,

$$\left[\left(f - y'\,\frac{\partial f}{\partial y'}\right) + g'\,\frac{\partial f}{\partial y'}\right]_{x\,=\,b} \;=\; 0\,, \tag{9.82}$$

gives

$$\left[\left(f - y'\,\frac{y'f}{1+y'^2}\right) + g'\,\frac{y'f}{1+y'^2}\right]_{x\,=\,b} \;=\; 0\,. \tag{9.83}$$

Upon simplifying this transversality condition, we find that

$$\left[\frac{f(1+y'g')}{1+y'^2}\right]_{x\,=\,b} \;=\; 0 \tag{9.84}$$

and that

$$y'(b) \;=\; -\frac{1}{g'(b)}\,, \tag{9.85}$$

where $y'(b)$ is the slope of the extremal at the terminal endpoint and $g'(b)$ is the slope of the boundary curve at its intersection with the extremal. Our last equation implies that these two curves must be perpendicular or orthogonal at this intersection. This geometric condition is enough to determine the terminal condition.

**Example 9.5.**

Let us find the extremum for the functional

$$J[y] \;=\; \int_a^b (1 \,+\, y'^2)\; dx \tag{9.86}$$

subject to the initial condition

$$y(0) \;=\; 0 \tag{9.87}$$

and the terminal condition that the extremal must intersect

$$y \;=\; \frac{1}{x} \tag{9.88}$$

for some, as yet, undetermined $x = b$.

The extrema satisfy

$$y \;=\; mx + c \tag{9.89}$$

and the first boundary condition implies that this equation simplifies to

$$y \;=\; mx\,. \tag{9.90}$$

The transversality condition is now

$$\left[\left(f - y'\frac{\partial f}{\partial y'}\right)\delta x + \frac{\partial f}{\partial y'}\,\delta y\right]_{x\,=\,b} \;=\; 0 \tag{9.91}$$

or

$$\left[(1 - y'^2)\delta x + 2\,y'\,\delta y\right]_{x\,=\,b} \;=\; 0\,. \tag{9.92}$$

Along the right boundary curve,

$$\delta y \;\approx\; -\frac{1}{x^2}\,\delta x \tag{9.93}$$

so that

$$\left[(1 - y'^2)\,\delta x - \frac{2\,y'}{x^2}\,\delta x\right]_{x\,=\,b} \;=\; 0\,. \tag{9.94}$$

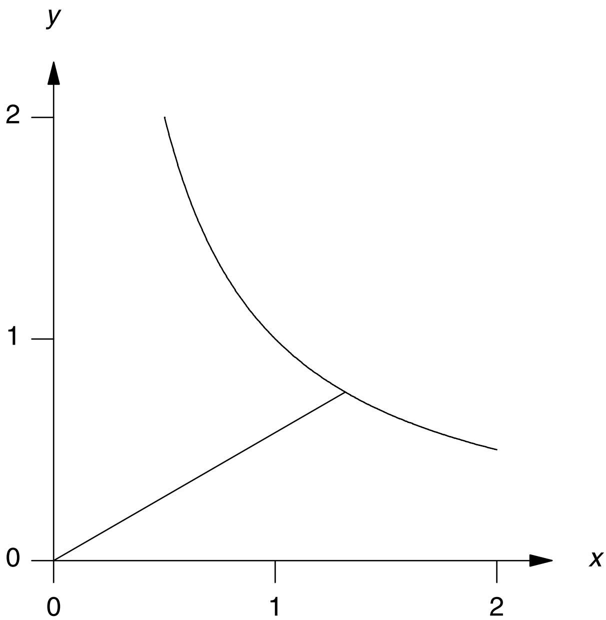



**Figure 9.3.** Transversal intersection

This reduces to the transversality condition

$$2\frac{m}{b^2} + m^2 - 1 \;=\; 0 \tag{9.95}$$

for our extremals.

Since the extremal and the boundary curve must intersect at $x = b$, it follows that

$$mb \;=\; \frac{1}{b} \tag{9.96}$$

or

$$m \;=\; \frac{1}{b^2}\,. \tag{9.97}$$

This last equation and the transversality condition, together, imply that

$$3m^2 - 1 \;=\; 0 \tag{9.98}$$

or

$$m = \frac{1}{\sqrt{3}}. \tag{9.99}$$

It also follows that

$$b = (3)^{1/4}. \tag{9.100}$$

In this case, the extremal and right boundary curve are not orthogonal at their intersection (see Figure 9.3). If they were, the intersection would be at (1, 1). However, they are still transversal.

## 9.3. Focal points

Conjugate points for fixed-endpoint problems give way to *focal points* for variable-endpoint problems.

The general solution,

$$y = \hat{y}(x, \alpha, \beta), \tag{9.101}$$

to the Euler–Lagrange equation has two constants of integration. For a fixed-endpoint problem, this two-parameter family of extremals simplifies to a one-parameter family of extremals that satisfy the left boundary condition. This one-parameter family may have an envelope. The conjugate point for an extremal is the point of contact between the extremal and its envelope.

For a variable-endpoint problem, the extremals that satisfy a transversality condition and that emanate out of a boundary curve (see Figure 9.4) may also possess an envelope. The focal point for an extremal is the point of contact between the extremal and this envelope. An extremal will not minimize or maximize a functional if the abscissa of the focal point lies within the interval of integration.

## 9.4. Case study: Neile's parabola

To find the shortest distance from a point, $(b, y_b)$, lying above the parabola $y = x^2$ to that parabola, we need to consider the functional

$$J[y] = \int_a^b \sqrt{1 + y'^2}\, dx \tag{9.102}$$

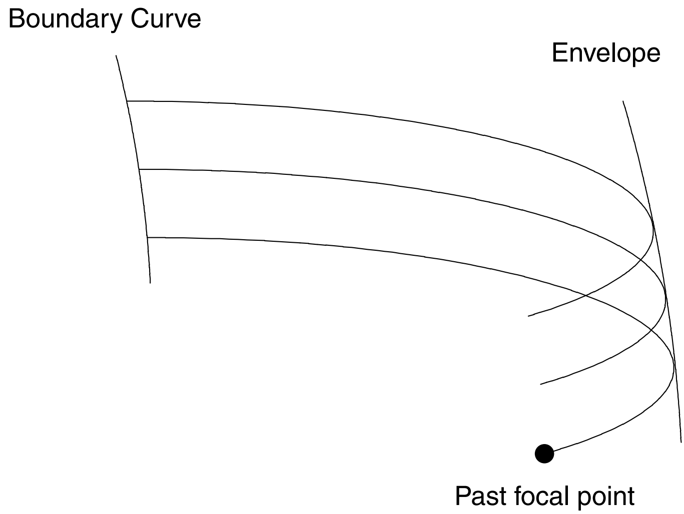



**Figure 9.4.** Envelope of focal points

along with the boundary conditions

$$y(a) = a^2, \quad y(b) = y_b. \tag{9.103}$$

The extremals for this problem are straight lines. Since this is a Fermat-type integral, transversality reduces to orthogonality. Here is a one-parameter family of lines that are orthogonal to the parabola:

$$y = -\frac{x}{2c} + \left(\frac{1}{2} + c^2\right). \tag{9.104}$$

This line intersects the parabola at a point,

$$x = c, \quad y = c^2, \tag{9.105}$$

where the slope of the parabola is $2c$.

Does this family of extremals have an envelope? Yes. Let's compute it. The c-discriminant is determined by the two equations

$$y = -\frac{x}{2c} + \left(\frac{1}{2} + c^2\right) \tag{9.106}$$

and

$$\frac{\partial y}{\partial c} = \frac{x}{2c^2} + 2\,c = 0. \tag{9.107}$$

It now follows that our envelope is given, parametrically, as

$$x = -4\,c^3\,, \quad y = 3\,c^2 + \frac{1}{2} \tag{9.108}$$

or, in explicit form, as

$$y = \frac{3}{4}(2x)^{2/3} + \frac{1}{2}\,. \tag{9.109}$$

This envelope is known as Neile's semicubical parabola (see Figure 9.5). This evolute of a parabola was discovered by William Neile in 1659. It was the first nontrivial algebraic curve to be rectified (to have its arc length calculated).

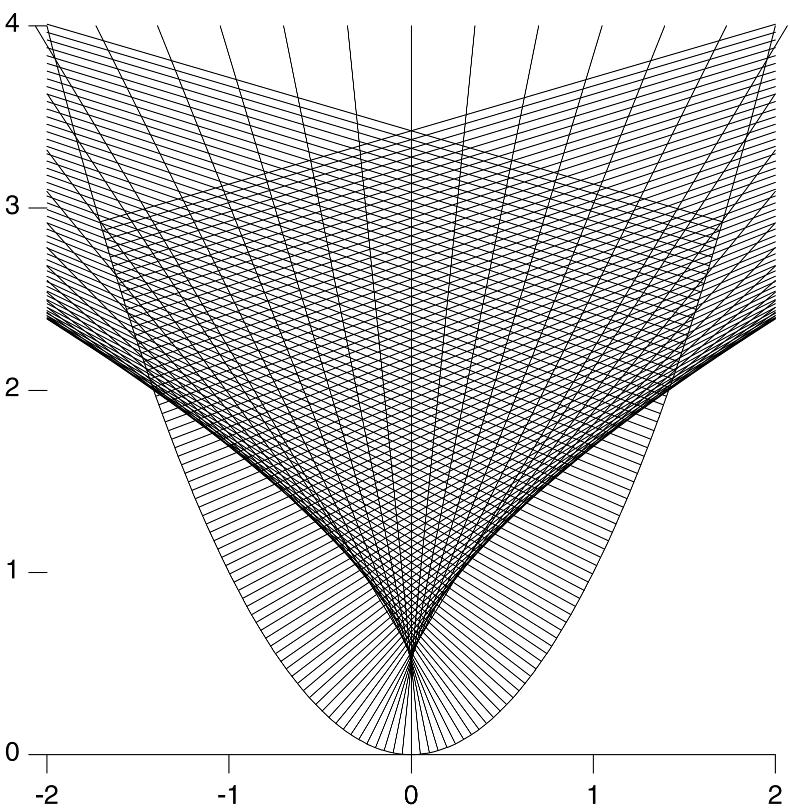



**Figure 9.5.** Neile's semicubical parabola

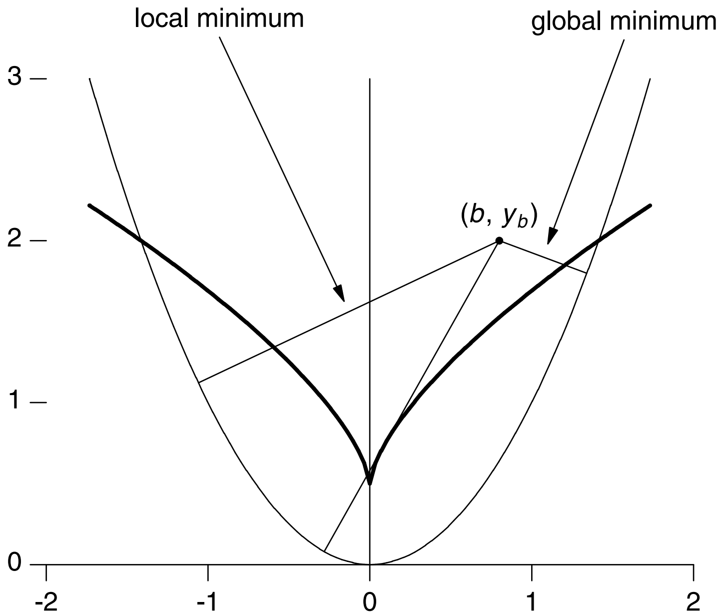



**Figure 9.6.** Local and global minima

Looking at Figure 9.5, we see that there are three extremals passing through each point above the evolute and only one extremal passing through each point below the evolute. We expect that a straight-line extremal will be a weak relative minimum if it does not pass through a focal point on its way to the point $(b, y_b)$, but that it will fail to be a minimum if it does pass through its focal point. Above the evolute, two extremals are relative minima while one is not. Figure 9.6 illustrates the situation.

We can think of this picture as describing the strategies that a sailor on Lake Michigan might consider. There is a closest point on the near shore (a global minimum), a closest point on the far shore (a local minimum), and a spurious extremal that leads to Gary, Indiana.

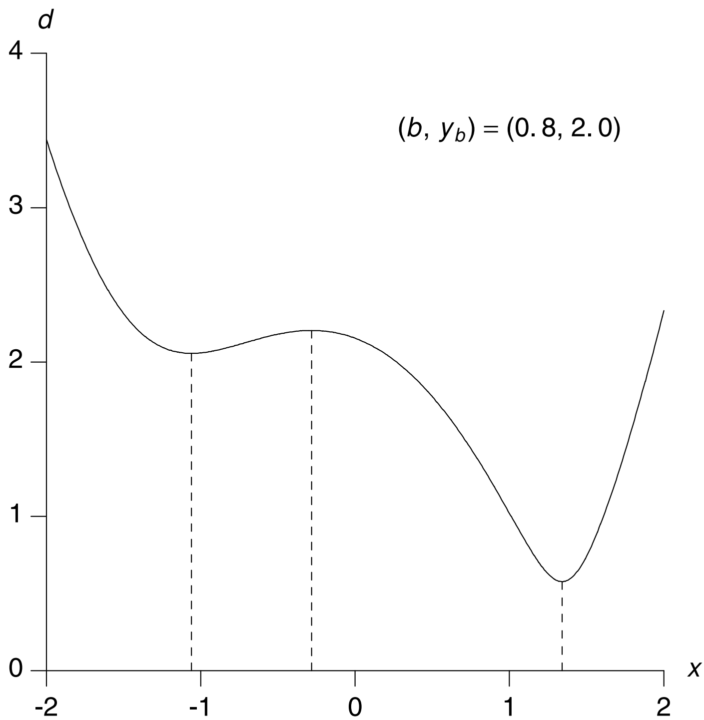



**Figure 9.7.** Distances from shore to $(0.8, 2.0)$

It is easy to verify these conclusions by evaluating the function

$$d(x) \;=\; \sqrt{(x-b)^2 + (x^2 - y_b)^2} \tag{9.110}$$

for the distance from the point $(b, y_b)$ to a point on the parabola, $(x, x^2)$, for various values of $x$.

For $(b, y_b) = (0.8, 2.0)$ (see Figure 9.7), $c = 1.341$ is the abscissa of the global minimum on the near shore and $c = -1.0595$ is the abscissa of the local minimum on the far shore. The extremal corresponding to $c = -0.2815$ clearly fails to produce a minimum. Indeed it seems to produce a local maximum, but don't take this too seriously; there are many curves other than straight lines that give longer distances.

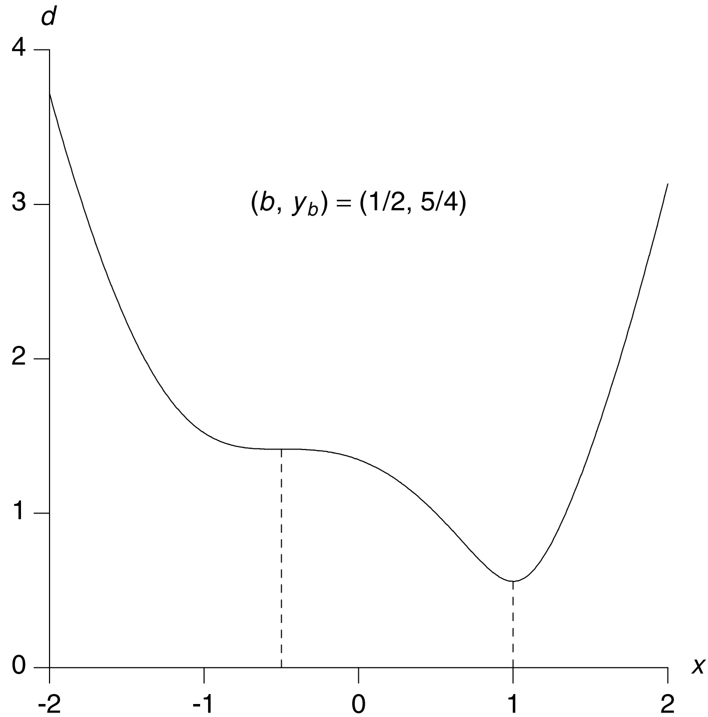



**Figure 9.8.** Distances from shore to $(1/2, 5/4)$

For a point right on the envelope, say for at $(0.5, 1.25)$ (see Figure 9.8), there is global minimum for $c = 1$ but no other relative minimum. The plot of distance does, however, have an inflection point.

## 9.5. Recommended reading

We borrow heavily in this chapter from the examples and the discussion in Sagan (1969). See Ebbinghaus (2007) and Carathéodory (2002) for more on Zermelo's navigation problem.

Mertens and Mingramm (2008) examine the brachistochrone problem with variable endpoints. They look for the fastest curve of descent between a point and a given curve or between two given curves. Smith (1974) uses a natural boundary condition to help design a thrilling amusement park chute-the-chute. Edelen (1981) reviews the proper use of transversality conditions in elastostatics.

Merrill (1919) considers necessary and sufficient conditions for an extremum for an isoperimetric problem with variable endpoints.

## 9.6. Exercises

**9.6.1. Steering angle.** Determine the steering strategy

$$\theta = \theta(x) \tag{9.111}$$

for Zermelo's problem with the parabolic velocity profile $v = x(1-x)$. Is your answer surprising? Why or why not?

**9.6.2. The brachistochrone to a vertical line.** Find the curve that minimizes the travel time of a heavy particle that starts at rest at the origin and that moves, under the force of gravity, to a given vertical line $x = b$. Assume that all points are in the same plane.

**9.6.3. Shortest distance.** Use the calculus of variations to find the shortest distance from the origin, $(x, y) = (0, 0)$, to the circle

$$(x-1)^2 + (y-2)^2 = 1\,. \tag{9.112}$$

**9.6.4. From a line to a circle.** Use transversality conditions at both ends to find the shortest distance between the line

$$y = x \tag{9.113}$$

and the circle

$$x^2 + (y-3)^2 = 1\,. \tag{9.114}$$

**9.6.5. From a parabola to a line.** Find the shortest distance between the parabola

$$y = x^2 \tag{9.115}$$

and the straight line

$$x - y = 5\,. \tag{9.116}$$

**9.6.6. A transversality condition.** Find the extremals for the functional

$$J[y] \;=\; \int\limits_0^b \frac{\sqrt{1+y'^2}}{y}\,dx \qquad (9.117)$$

subject to the boundary condition

$$y(0) \;=\; 0 \qquad (9.118)$$

and to the condition that the right endpoint, $(b, y_b)$, can move along the circumference of the circle

$$(x-9)^2 + y^2 \;=\; 9\,. \qquad (9.119)$$

Chapter 10

# Broken Extremals

## 10.1. The Weierstrass–Erdmann corner conditions

So far, we have only dealt with continuously differentiable extremals. Our transversality conditions allow us to do better.

**Example 10.1.**

Consider the functional

$$J[y] \;=\; \int_{-1}^{+1} y^2\,(1 - y')^2\;dx \tag{10.1}$$

subject to the boundary conditions

$$y(-1) \;=\; 0\,, \quad y(+1) \;=\; 1\,. \tag{10.2}$$

The independent variable $x$ does not appear in the integrand. The corresponding first integral quickly reduces to

$$y^2\,(1 - {y'}^2) \;=\; \alpha\,. \tag{10.3}$$

It looks like

$$y' \;=\; 0\,, \quad y \;=\; \sqrt{\alpha} \tag{10.4}$$

is a solution, but this is a spurious solution of the first integral unless $y = 0$. More generally, the solutions are hyperbolas. These hyperbolas include the degenerate hyperbola $y = x$. Let us key in on the two solutions

$$y = 0\,, \quad y = x\,. \tag{10.5}$$

Both of these two solutions make the functional zero. You cannot do better, since the integral is clearly nonnegative. Unfortunately, neither of these two solutions satisfies the boundary conditions.

Let us instead consider the function

$$y = \begin{cases} 0\,, & -1 \le x \le 0\,, \\ x\,, & 0 \le x \le 1\,. \end{cases} \tag{10.6}$$

This function satisfies the boundary conditions. It also causes the integral to vanish. This function is not, however, a continuously differentiable solution of the Euler–Lagrange equation. Each *section* of this function *is*, however, a continuously differentiable solution of the Euler–Lagrange equation.

The optimal solution in the above example was *piecewise* continuously differentiable rather than continuously differentiable. In other words, it had a corner. In the above example, the corner was at the origin. In general, we need to locate a corner $(c, y_c)$ for the functional

$$J[y] = \int_a^b f(x, y, y')\, dx \tag{10.7}$$

from first principles. To do this, we will rewrite $J[y]$ as the sum of two functionals,

$$\begin{aligned} J[y] &= J_1[y] + J_2[y] \\ &= \int_a^c f(x, y, y')\, dx + \int_c^b f(x, y, y')\, dx\,, \end{aligned} \tag{10.8}$$

so that we can break the first variation, $\delta J$, into two components,

$$\delta J = \delta J_1 + \delta J_2\,. \tag{10.9}$$

The endpoints at $a$ and $b$ are fixed. We require that $y(x)$ be continuous (but not continuously differentiable) at $c$. The curves

$y(x)$ must be extremals on the segments $[a, c]$ and $[c, b]$, and so the contribution to the first variation from the interior of each interval vanishes. That leaves the two contributions that arise at $x = c$ from varying the right endpoint of $J_1[y]$ and the left endpoint of $J_2[y]$.

Using transversality conditions, we can write these contributions as

$$\delta J_1 \;=\; \left[\left(f - y'\,\frac{\partial f}{\partial y'}\right)\delta x + \frac{\partial f}{\partial y'}\,\delta y\right]_{x\,=\,c\,-\,0} \tag{10.10}$$

and

$$\delta J_2 \;=\; -\left[\left(f - y'\,\frac{\partial f}{\partial y'}\right)\delta x + \frac{\partial f}{\partial y'}\,\delta y\right]_{x\,=\,c\,+\,0}\,. \tag{10.11}$$

Here, the subscript $x = c - 0$ indicates that we need to take the limit (of the expression in parentheses) as $x$ approaches $c$ from the left. The subscript $x = c + 0$, in turn, signifies that we need to take the limit from the right.

We require, as usual, that

$$\delta J \;=\; \delta J_1 + \delta J_2 \;=\; 0\,. \tag{10.12}$$

It follows that

$$\left(f - y'\,\frac{\partial f}{\partial y'}\right)\Bigg|_{x\,=\,c+0}^{x\,=\,c-0}\delta c + \frac{\partial f}{\partial y'}\Bigg|_{x\,=\,c+0}^{x\,=\,c-0}\delta y_c \;=\; 0\,. \tag{10.13}$$

Since $\delta c$ and $\delta y_c$ are arbitrary, we must now impose the conditions

$$\frac{\partial f}{\partial y'}\bigg|_{x\,=\,c-0} \;=\; \frac{\partial f}{\partial y'}\bigg|_{x\,=\,c+0}\,, \tag{10.14}$$

$$\left(f - y'\,\frac{\partial f}{\partial y'}\right)_{x\,=\,c-0} \;=\; \left(f - y'\,\frac{\partial f}{\partial y'}\right)_{x\,=\,c+0}\,. \tag{10.15}$$

These two conditions are known as the *Weierstrass–Erdmann (corner) conditions.*

Let us tally constants of integration. We have two extremals, each with two constants of integration, for a total of four constants of integration. The boundary conditions at $x = a$ and $x = b$ determine two of these four constants. The two Weierstrass–Erdmann conditions determine the other two constants.

**Example 10.2.**

Let us revisit the functional

$$J[y] \;=\; \int_{-1}^{+1} f(x, y, y')\; dx \;=\; \int_{-1}^{+1} y^2 (1 - y')^2\; dx \tag{10.16}$$

with boundary conditions

$$y(-1) \;=\; 0\,, \quad y(+1) \;=\; 1\,. \tag{10.17}$$

The first integral

$$f - y' \frac{\partial f}{\partial y'} \;=\; \alpha \tag{10.18}$$

implies that

$$y^2 (1 - {y'}^2) \;=\; \alpha\,. \tag{10.19}$$

The trivial solution,

$$y \;=\; 0\,, \tag{10.20}$$

is a true solution of the full Euler–Lagrange equation. For $y \neq 0$,

$$y' \;=\; \pm \sqrt{1 - \frac{\alpha}{y^2}}\,. \tag{10.21}$$

We thus have several possible solution curves. At corners, we switch from one solution curve to another.

Let us now look at

$$\frac{\partial f}{\partial y'} \;=\; -2y^2 (1 - y') \tag{10.22}$$

and

$$f - y' \frac{\partial f}{\partial y'} \;=\; y^2 (1 - {y'}^2)\,. \tag{10.23}$$

Since $y(x)$ is continuous, it follows, from the first Weierstrass–Erdmann corner condition, that $y'(x)$ is also continuous, *unless* $y_c = y(c) = 0$. For $y(c) = 0$, $y'$ may be discontinuous with $y = 0$ to the left of the corner and $y' = 1$ to the right of the corner. This combination

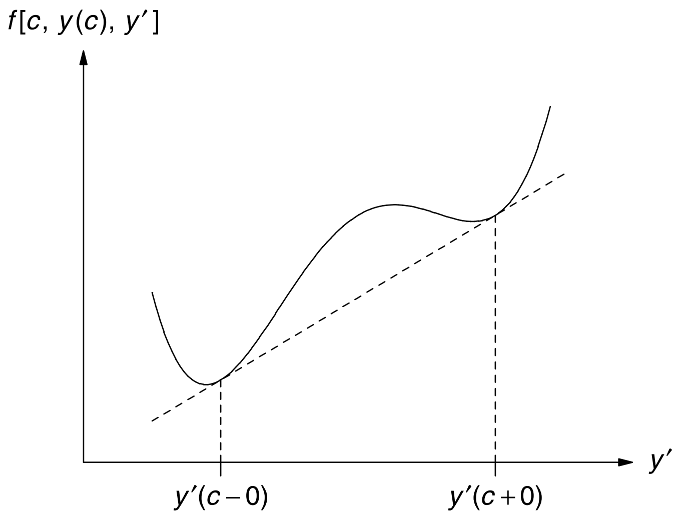



**Figure 10.1.** Indicatrix

of values will also satisfy the second Weierstrass–Erdmann corner condition. Thus, the broken solution,

$$y = \begin{cases} 0, & -1 \le x \le 0, \\ x, & 0 \le x \le 1, \end{cases} \tag{10.24}$$

is consistent with the Weierstrass–Erdmann corner conditions.

The Weierstrass–Erdmann corner conditions have a simple geometric interpretation. If you plot $f(x, y, y')$ as a function of $y'$, for fixed values of $x$ and $y$, you get a curve known as the *characteristic* (Sagan, 1969) or *indicatrix* (Akhiezer, 1962; Gelfand and Fomin, 1963). See Figure 10.1. The first corner condition tells us that the slopes of the tangents to the indicatrix at the points $y'(c - 0)$ and $y'(c + 0)$ must be the same,

$$\frac{\partial f}{\partial y'}(c, y(c), y'(c-0)) = \frac{\partial f}{\partial y'}(c, y(c), y'(c+0)) . \tag{10.25}$$

We will call this common slope $p(c)$. Since the two tangents have the same slope, they must be parallel. The second corner condition,

$$\left(f - y'\,\frac{\partial f}{\partial y'}\right)_{x\,=\,c-0} \;=\; \left(f - y'\,\frac{\partial f}{\partial y'}\right)_{x\,=\,c+0}, \qquad (10.26)$$

can be rewritten as

$$\begin{aligned} f(c, y(c), y'(c+0)) \;&=\; f(c, y(c), y'(c-0)) \\ &\quad + p(c)\,\left[y'(c+0) - y'(c-0)\right]. \end{aligned} \qquad (10.27)$$

It now follows that the two tangents must be the same line.

We see the following:

(1) a corner can occur at $(x, y) = [c, y(c)]$ only if the corresponding indicatrix has a tangent that touches two or more points and that

(2) a corner is immediately ruled out if $f$ is strictly convex,

$$\frac{\partial^2 f}{\partial y'^2}(c, y(c), y') \;>\; 0, \qquad (10.28)$$

or concave,

$$\frac{\partial^2 f}{\partial y'^2}(c, y(c), y') \;<\; 0, \qquad (10.29)$$

(for all $y'$).

**Example 10.3.**

The integrand

$$f(x, y, y') \;=\; y^2(1 - y')^2 \qquad (10.30)$$

has two, qualitatively different, characteristics (see Figure 10.2). For $y \neq 0$, the characteristic is strictly convex. For $y = 0$, the characteristic is a horizontal line. Only in the latter case can we satisfy the Weierstrass–Erdmann corner conditions.

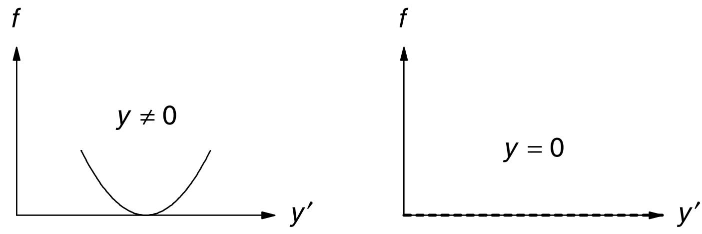



**Figure 10.2.** Characteristics for $y^2(1-y')^2$

**Example 10.4.**

For the integrand

$$f(x,y,y') \;=\; y\sqrt{1+y'^2} \tag{10.31}$$

it is easy to show that

$$f_{y'y'} \;=\; \frac{y}{(1+y'^2)^{3/2}}\,. \tag{10.32}$$

It follows that a minimal surface of revolution cannot have a corner unless $y=0$.

**Example 10.5.**

Consider the functional

$$J[y] \;=\; \int_0^2 f(x,y,y')\,dx \;=\; \int_0^2 (y'+1)^2 y'^2\,dx \tag{10.33}$$

with the boundary conditions

$$y(0) \;=\; 1\,, \quad y(2) \;=\; 0\,. \tag{10.34}$$

Let us approach this problem one step at a time. What are the extremals? Since the integrand is missing explicit dependence on both $x$ and $y$, there are two obvious first integrals. Since $y$ is a cyclic or ignorable coordinate,

$$\frac{\partial f}{\partial y'} \;=\; 4\,{y'}^3 + 6\,{y'}^2 + 2\,y' \;=\; c\,. \tag{10.35}$$

It follows that $y'$ must be a constant and that

$$y = mx + b\,. \tag{10.36}$$

Applying the boundary conditions, we obtain the continuously differentiable solution

$$\hat{y}(x) = -\frac{1}{2}\,x + 1\,. \tag{10.37}$$

How about solutions with corners? Since the integrand $f$ is independent of $x$ and $y$, we get a single characteristic for all choices of $x$ and $y$. Plotting the indicatrix (see Figure 10.3), we see that we have one line of double tangency with

$$y' = -1 \quad \text{and} \quad y' = 0 \tag{10.38}$$

at the points of tangency. Moreover, this occurs for *all* choices of $x$ and $y$. Indeed, since

$$\begin{aligned} \frac{\partial f}{\partial y'} &= 4\,{y'}^3 + 6\,{y'}^2 + 2\,y' \\ &= 2\,y'(2y' + 1)(y' + 1) \end{aligned} \tag{10.39}$$

and

$$\begin{aligned} f - y'\,\frac{\partial f}{\partial y'} &= -(3{y'}^4 + 4{y'}^3 + {y'}^2) \\ &= -{y'}^2(3y' + 1)(y' + 1)\,, \end{aligned} \tag{10.40}$$

it is clear that choosing $y' = 0$ from one direction and $y' = -1$ from the other will keep both of these two expressions continuous. Every piecewise continuously differentiable solution must, therefore, be composed of straight-line segments making the angles 0 or $-\pi/4$ with the positive $x$-axis.

There are two broken extremals with one corner (see Figure 10.4). There are, however, an *infinite* number of solutions with more than one corner. All of the broken extremals reduce the definite integral to zero; they are (improper) absolute minima.

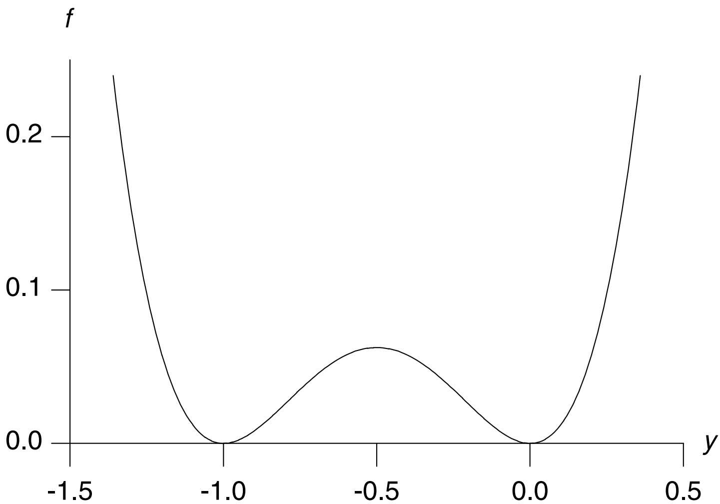



**Figure 10.3.** Characteristic for $(y' + 1)^2\, {y'}^2$

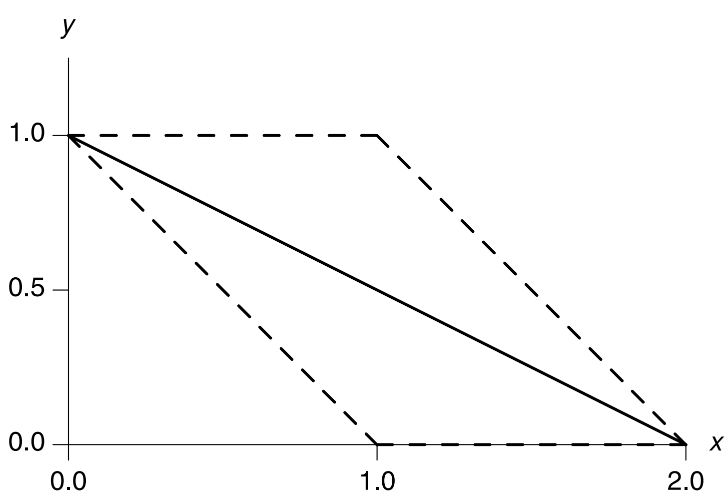



**Figure 10.4.** Broken extremals

## 10.2. Carathéodory's indicatrix

Let us return to the homogeneous problem. Remember that we first rewrote the functional

$$J[y] \;=\; \int_a^b f(x,y,y')\;dx\,, \tag{10.41}$$

for the ordinary problem, as

$$J[\gamma] \;=\; \int_{t_a}^{t_b} F\,(x,y,\dot{x},\dot{y})\;dt\,, \tag{10.42}$$

for the homogeneous or parametric problem, with

$$F(x,y,\dot{x},\dot{y}) \;\equiv\; f\left(x(t),y(t),\frac{\dot{y}(t)}{\dot{x}(t)}\right)\dot{x}(t)\,. \tag{10.43}$$

For this homogeneous problem, it is easy to show that the two corner conditions are

$$\left.\frac{\partial F}{\partial \dot{x}}\right|_{t\,=\,t_c-0} \;=\; \left.\frac{\partial F}{\partial \dot{x}}\right|_{t\,=\,t_c+0}\,, \tag{10.44}$$

$$\left.\frac{\partial F}{\partial \dot{y}}\right|_{t\,=\,t_c-0} \;=\; \left.\frac{\partial F}{\partial \dot{y}}\right|_{t\,=\,t_c+0}\,. \tag{10.45}$$

What is the parametric analog of the characteristic?

To answer that question, we will need to assume that $F$ is positive definite, with

$$F(x,y,\dot{x},\dot{y}) \;>\; 0 \tag{10.46}$$

for all

$$\dot{x}^2+\dot{y}^2 \;>\; 0 \tag{10.47}$$

in some neighborhood of a given $x$ and $y$. This is a restriction on what integrands we may consider, but many homogeneous problems satisfy this condition.

Let us now look at a level curve of $F$ in the $(\dot{x},\dot{y})$ plane. The curve

$$F(x_0,y_0,\dot{x},\dot{y}) \;=\; 1 \tag{10.48}$$

is called *Carathéodory's indicatrix* at $(x_0, y_0)$. Why did we choose this particular level curve? Well, $F$ is homogeneous in $\dot{x}$ and $\dot{y}$. Any other level curve

$$F(x_0, y_0, \dot{x}, \dot{y}) \;=\; c \tag{10.49}$$

can be rewritten as

$$F\left(x_0, y_0, \frac{\dot{x}}{c}, \frac{\dot{y}}{c}\right) \;=\; 1 \tag{10.50}$$

and can thus be thought of as a rescaled version of Carathéodory's indicatrix. The size of level curves of $F$ can vary, but their shape is always the same. In that sense, Carathéodory's indicatrix is as good as any other level curve.

Let us now introduce polar coordinates,

$$\dot{x} \;=\; r\,\cos\,\theta\,, \quad \dot{y} \;=\; r\,\sin\,\theta\,. \tag{10.51}$$

You should note that

$$y' \;=\; \frac{\dot{y}}{\dot{x}} \;=\; \tan\,\theta \tag{10.52}$$

and that we may obtain all possible slopes $y'$ by varying $\theta$. With polar coordinates, Carathéodory's indicatrix now takes the form

$$F(x_0, y_0, r\,\cos\,\theta, r\,\sin\,\theta) \;=\; 1\,. \tag{10.53}$$

Since $F$ is positively homogeneous of degree one in its derivatives, we can factor out $r$,

$$r\,F(x_0, y_0, \cos\,\theta, \sin\,\theta) \;=\; 1\,. \tag{10.54}$$

It now follows that our level curve can be written in the polar form

$$r \;=\; \frac{1}{F(x_0, y_0, \cos\,\theta, \sin\,\theta)}\,. \tag{10.55}$$

Since the trigonometric functions in $F$ are $2\pi$-periodic functions, Carathéodory's indicatrix is a closed curve. It is easy to see that this closed curve contains the origin. We shall soon think of the angles $\theta$ (and slopes $y'$) for points on Carathéodory's indicatrix as the angles (and slopes) of the endpoints of extremals.

**Example 10.6.**

Let

$$F(x, y, \dot{x}, \dot{y}) = k(x, y)\sqrt{\dot{x}^2 + \dot{y}^2}\,, \tag{10.56}$$

for

$$k(x, y) > 0\,. \tag{10.57}$$

Carathéodory's indicatrix,

$$F(x_0, y_0, \dot{x}, \dot{y}) = 1\,, \tag{10.58}$$

is, in this case,

$$k(x_0, y_0)\sqrt{\dot{x}^2 + \dot{y}^2} = 1 \tag{10.59}$$

or

$$\dot{x}^2 + \dot{y}^2 = \frac{1}{k^2(x_0, y_0)}\,. \tag{10.60}$$

This is a circle for every point $(x_0, y_0)$. In polar form,

$$r = \frac{1}{F(x_0, y_0, \cos\theta, \sin\theta)} = \frac{1}{k(x_0, y_0)}\,. \tag{10.61}$$

For the characteristic or indicatrix, a corner corresponded to a double tangency in $y'$. A corner was possible only if a line was tangent to the characteristic for two different values of $y'$. For Carathéodory's indicatrix, a corner corresponds to a double tangency in $\theta$. The basic idea is simple, but getting there takes a bit of effort.

Carathéodory's indicatrix is a level curve of $F$ in the $(\dot{x}, \dot{y})$ plane. In this plane,

$$\nabla F = \left(\frac{\partial F}{\partial \dot{x}}, \frac{\partial F}{\partial \dot{y}}\right)\,. \tag{10.62}$$

This vector is orthogonal to the indicatrix at a point, $(\dot{x}, \dot{y})$, of the indicatrix.

Suppose that we have a variable-endpoint problem and that we specify that the extremal must terminate on the curve

$$x = \phi(u)\,, \quad y = \psi(u)\,. \tag{10.63}$$

Suppose that this occurs at

$$x_0 = \phi(u_0)\,, \quad y_0 = \psi(u_0)\,. \tag{10.64}$$

Then the transversality condition reads

$$\frac{\partial F}{\partial \dot{x}}\,\phi'(u_0) + \frac{\partial F}{\partial \dot{y}}\,\psi'(u_0) \;=\; 0\,. \tag{10.65}$$

This may be rewritten as

$$\left(\frac{\partial F}{\partial \dot{x}}, \frac{\partial F}{\partial \dot{y}}\right) \cdot (\phi'(u_0), \psi'(u_0)) \;=\; 0 \tag{10.66}$$

so that

$$\left(\frac{\partial F}{\partial \dot{x}}, \frac{\partial F}{\partial \dot{y}}\right) \perp (\phi'(u_0), \psi'(u_0))\;\;. \tag{10.67}$$

Since $\nabla F$ is also orthogonal to the tangent line to the indicatrix at a given point, $(\dot{x}, \dot{y})$, of the indicatrix, it follows that

$$(\phi'(u_0), \psi'(u_0)) \tag{10.68}$$

is parallel to a tangent line of the indicatrix. This tangent line determines, at its point of tangency, the angle $\theta$ and the slope $y'$ of the extremal at its intersection with the boundary curve.

How about the geometric relationship between Carathéodory's indicatrix and our corner conditions? Let us work through this carefully since this is what we are really after. The tangent line to the indicatrix at the point $(\dot{x}_0, \dot{y}_0)$ satisfies

$$(\dot{x} - \dot{x}_0, \dot{y} - \dot{y}_0) \cdot [F_{\dot{x}}(x_0, y_{0,}\dot{x}_0, \dot{y}_0), F_{\dot{y}}(x_0, y_{0,}\dot{x}_0, \dot{y}_0)] \;=\; 0 \tag{10.69}$$

or

$$(\dot{x} - \dot{x}_0)F_{\dot{x}}(x_0, y_{0,}\dot{x}_0, \dot{y}_0) + (\dot{y} - \dot{y}_0)F_{\dot{y}}(x_0, y_{0,}\dot{x}_0, \dot{y}_0) \;=\; 0\,. \tag{10.70}$$

We can simplify this last equation. Since positively homogeneous functions such as $F$ satisfy Euler's identity,

$$\dot{x}\,F_{\dot{x}}(x, y, \dot{x}, \dot{y}) + \dot{y}\,F_{\dot{y}}(x, y, \dot{x}, \dot{y}) \;=\; F(x, y, \dot{x}, \dot{y})\,, \tag{10.71}$$

we may write

$$\begin{aligned} &\dot{x}_0\,F_{\dot{x}}(x_0, y_0, \dot{x}_0, \dot{y}_0) + \dot{y}_0\,F_{\dot{y}}(x_0, y_0, \dot{x}_0, \dot{y}_0) \\ &\quad = \; F(x_0, y_0, \dot{x}_0, \dot{y}_0)\,. \end{aligned} \tag{10.72}$$

But, $(\dot{x}_0, \dot{y}_0)$ is a point on the indicatrix, so that

$$F(x_0, y_0, \dot{x}_0, \dot{y}_0) \;=\; 1\,. \tag{10.73}$$

As a result, the equation of our tangent line simplifies to

$$\dot{x}\, F_{\dot{x}}(x_0, y_0, \dot{x}_0, \dot{y}_0) + \dot{y}\, F_{\dot{y}}(x_0, y_0, \dot{x}_0, \dot{y}_0) \;=\; 1\,. \tag{10.74}$$

Let us introduce one other useful fact. $F_{\dot{x}}$ and $F_{\dot{y}}$ are homogeneous of degree zero in their derivatives. To see this, start with the positive homogeneity of $F$,

$$F(x, y, \lambda\dot{x}, \lambda\dot{y}) \;=\; \lambda\, F(x, y, \dot{x}, \dot{y}) \tag{10.75}$$

and differentiate with respect to $\dot{x}$,

$$F_{\dot{x}}(x, y,\, \lambda\dot{x}, \lambda\dot{y})\, \lambda \;=\; \lambda\, F_{\dot{x}}(x, y, \dot{x}, \dot{y})\,. \tag{10.76}$$

Canceling a $\lambda$ on each side, we get

$$F_{\dot{x}}(x, y, \lambda\dot{x}, \lambda\dot{y}) \;=\; F_{\dot{x}}(x, y, \dot{x}, \dot{y})\,, \tag{10.77}$$

our desired result. Proceed analogously for $F_{\dot{y}}$. The homogeneity of degree zero of $F_{\dot{x}}$ and $F_{\dot{y}}$ in their derivatives means that we can always write

$$F_{\dot{x}}(x, y, r\cos\theta, r\sin\theta) \;=\; F_{\dot{x}}(x, y, \cos\theta, \sin\theta)\,, \tag{10.78}$$

$$F_{\dot{y}}(x, y, r\cos\theta, r\sin\theta) \;=\; F_{\dot{y}}(x, y, \cos\theta, \sin\theta)\,. \tag{10.79}$$

We may thus write the equation for the tangent line to our indicatrix, equation (10.74), as

$$\dot{x}\, F_{\dot{x}}(x_0, y_0, \cos\theta_0, \sin\theta_0) + \dot{y}\, F_{\dot{y}}(x_0, y_0, \cos\theta_0, \sin\theta_0) \;=\; 1\,. \tag{10.80}$$

Okay, we are now all set. Suppose that we have a corner at $(x_0, y_0)$, that the angle of an incoming extremal is $\theta_1$, and that the angle of the outgoing extremal is $\theta_2$. Then our corner conditions imply that

$$F_{\dot{x}}(x_0, y_0, \cos\theta_1, \sin\theta_1) \;=\; F_{\dot{x}}(x_0, y_0, \cos\theta_2, \sin\theta_2)\,, \tag{10.81}$$

$$F_{\dot{y}}(x_0, y_0, \cos\theta_1, \sin\theta_1) \;=\; F_{\dot{y}}(x_0, y_0, \cos\theta_2, \sin\theta_2)\,. \tag{10.82}$$

(Notice that we have used the fact that $F_{\dot{x}}$ and $F_{\dot{y}}$ are homogeneous of degree zero in their derivatives to eliminate the radii from the arguments $\dot{x}$ and $\dot{y}$.) Looking back at the equation (10.80) for the tangent line to our indicatrix, we see that the points on our indicatrix corresponding to angles $\theta_1$ and $\theta_2$ have the same tangent line. A corner shows up in Carathéodory's indicatrix as a line of double tangency.

**Example 10.7.**

Consider the functional

$$J[x(t), y(t)] \; = \; \int\limits_{(0,0)}^{(-1,0)} \frac{\dot{x}^2 + \dot{y}^2}{\sqrt{2\,(\dot{x}^2 + \dot{y}^2)} + \dot{x}}\, dt\,. \tag{10.83}$$

This is clearly a homogeneous problem since the integrand,

$$F(x, y, \dot{x}, \dot{y}) \; = \; \frac{\dot{x}^2 + \dot{y}^2}{\sqrt{2(\dot{x}^2 + \dot{y}^2)} + \dot{x}}\,, \tag{10.84}$$

is positively homogeneous of degree one in the derivatives.

One can show that the extremals for his problem are straight lines. One obvious extremal that satisfies the boundary conditions is

$$x \; = \; -t, \quad y \; = \; 0\,, \quad 0 \leq t \leq 1, \tag{10.85}$$

with

$$J[x(t), y(t)] \; = \; \int\limits_0^1 \frac{1}{\sqrt{2} - 1}\, dt \; = \; \sqrt{2} + 1\,. \tag{10.86}$$

Are there any other possible solutions, say with corners?

The polar form of the indicatrix,

$$r \; = \; \frac{1}{F(x, y, \cos\,\theta, \sin\,\theta)}\,, \tag{10.87}$$

gives us

$$r \; = \; \sqrt{2} + \cos\,\theta\,. \tag{10.88}$$

This is a special case of the limaçon (snail) of Pascal (see Figure 10.5).

This curve has a double vertical tangent when

$$r \; = \; \frac{b}{\cos\,\theta} \; = \; \sqrt{2} + \cos\,\theta \tag{10.89}$$

or

$$\cos^2\theta + \sqrt{2}\,\cos\,\theta - b \; = \; 0\,. \tag{10.90}$$

This occurs when

$$\cos\,\theta \; = \; \frac{-\sqrt{2} \pm \sqrt{2 + 4b}}{2}\,. \tag{10.91}$$

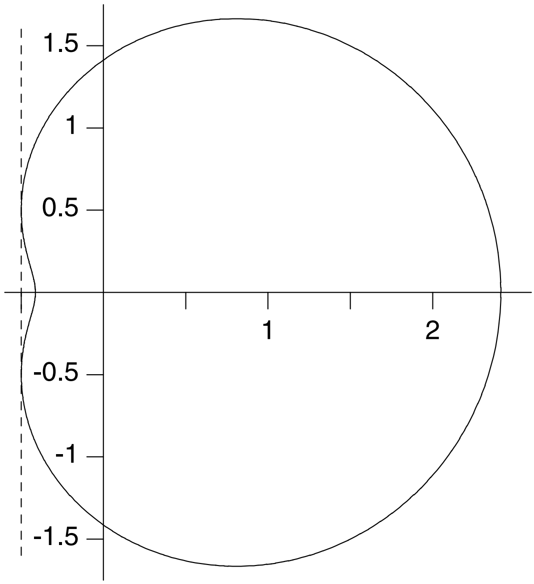



**Figure 10.5.** Pascal's snail

This equation has a double root when

$$b = -\frac{1}{2}, \tag{10.92}$$

which implies that

$$\cos\theta = -\frac{\sqrt{2}}{2}. \tag{10.93}$$

It now follows that

$$\theta = \frac{3\pi}{4}, \frac{5\pi}{4}. \tag{10.94}$$

A corner is thus possible at every point of the $(x, y)$ plane if the extremal goes into the corner at an angle of $3\pi/4$ relative to the positive $x$-axis and leaves the corner at an angle of $5\pi/4$ or if it goes into the corner at an angle of $5\pi/4$ and leaves the corner at an angle of $3\pi/4$.

Let us now look for broken extremals that connect $(-1, 0)$ and $(0, 0)$. These broken-extremal solutions are make up of line segments

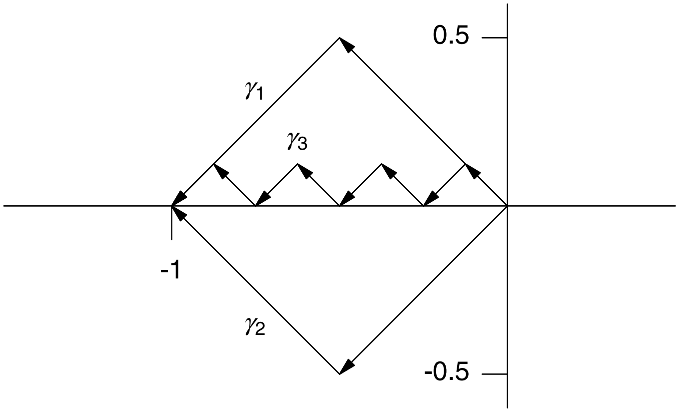



**Figure 10.6.** Broken-extremal solutions

that enter or leave corners at angles of $3\pi/4$ or $5\pi/4$ (see Figure 10.6).

For $\gamma_1$, with one corner at $(-0.5, 0.5)$,

$$x(t) \;=\; -t\,, \quad y(t) \;=\; t \tag{10.95}$$

for $0 \le t \le 1/2$, but

$$x(t) \;=\; -t\,, \quad y(t) \;=\; -t+1 \tag{10.96}$$

for $1/2 \le t \le 1$. Thus

$$J[\gamma_1] \;=\; \int_0^1 \frac{2}{2-1}\, dt \;=\; 2 \tag{10.97}$$

and we do better using a broken extremal rather than an unbroken extremal.

## 10.3. Recommended reading

Broken extremals, i.e., solutions to problems in the calculus of variations with corners, are also called discontinuous solutions or extremaloids. See Graves (1930a) for an early review paper on discontinuous solutions and Bolza (1908), Dresden (1908), and Graves

(1930b) for the extension of the Jacobi condition to curves with corners.

We borrowed heavily from the book by Sagan (1969) in our treatment of Carathéodory's indicatrix. For more on Carathéodory's indicatrix, see also Dresden (1907).

## 10.4. Exercises

**10.4.1. One corner.** Find a solution with one corner point for the problem of minimizing

$$\int_0^4 (y' - 1)^2 \, (y' + 1)^2 \; dx \tag{10.98}$$

subject to the boundary conditions

$$y(0) \;=\; 0\,, \quad y(4) \;=\; 2\,. \tag{10.99}$$

**10.4.2. Hunting corners.** Consider the functional

$$J[y] \;=\; \int_a^b (y'^2 + 2xy - y^2) \; dx \tag{10.100}$$

subject to the boundary conditions

$$y(a) \;=\; y_a\,, \quad y(b) \;=\; y_b\,. \tag{10.101}$$

What kind of corner solutions exist for this problem?

**10.4.3. At a loss for corners.** Consider the functional

$$J[y] \;=\; \int_a^b (y'^2 + xy' + x^2) \; dx \tag{10.102}$$

subject to the boundary conditions

$$y(a) \;=\; y_a\,, \quad y(b) \;=\; y_b\,. \tag{10.103}$$

Show that extremals must be smooth as a consequence of the corner conditions.

**10.4.4. Cutting corners.** Consider the functional

$$J[y] \;=\; \int_a^b \left(y' - 6{y'}^2\right)\,dx \tag{10.104}$$

subject to the boundary conditions

$$y(a) \;=\; y_a\,, \quad y(b) \;=\; y_b\,. \tag{10.105}$$

Use the corner conditions to determine whether this integral has extremals with corners.

Chapter 11

# Strong Variations

## 11.1. Troubles with weak variations

The Euler–Lagrange equation, the strengthened Legendre condition, and the strengthened Jacobi condition are sufficient conditions for a weak relative minimum. This was the highwater mark of classical (pre-Weierstrassian) calculus of variations. Unfortunately, these conditions are not, by themselves, sufficient for a strong relative minimum. The root of the problem can already be seen in a corner problem that we previously considered in Chapter 10.

**Example 11.1.**

Consider the functional

$$J[y] \;=\; \int_a^b (y' + 1)^2 {y'}^2 \, dx \tag{11.1}$$

with the boundary conditions

$$y(a) \;=\; y_a\,, \quad y(b) \;=\; y_b\,. \tag{11.2}$$

Since the dependent variable is missing, the Euler–Lagrange equation for this problem reduces to

$$\frac{\partial f}{\partial y'} \;=\; 4{y'}^3 + 6{y'}^2 + 2y' \;=\; \alpha\,. \tag{11.3}$$

It follows that $y'$ must be a constant and that the extremal for this problem is the straight line

$$y = mx + k \tag{11.4}$$

that connects the two boundary points.

Let's look at Legendre's test. For our extremal,

$$R = \frac{\partial^2 f}{\partial {y'}^2} = 2(6m^2 + 6m + 1) \tag{11.5}$$

(see Figure 11.1). Where is $R$ positive or negative? Let $m_1$ and $m_2$ be the two roots of the equation

$$6m^2 + 6m + 1 = 0\,. \tag{11.6}$$

Solving for the roots,

$$m_{1,2} = \frac{-6 \pm \sqrt{36 - 24}}{12} = \frac{-3 \pm \sqrt{3}}{6}\,, \tag{11.7}$$

we see that

$$m_1 = -0.788675\,, \quad m_2 = -0.21132\,. \tag{11.8}$$

Moreover, since $R$ is a concave-up function of $m$, it follows that $R > 0$ for $m < m_1$ or $m > m_2$, but that $R < 0$ for $m_1 < m < m_2$. Thus, the strengthened Legendre condition for a minimum is satisfied for $m < m_1$ or for $m > m_2$ and the corresponding condition for a maximum is satisfied for $m_1 < m < m_2$

For the Jacobi condition, it is sufficient to note that

$$u(x) = c_1 u_1(x) + c_2 u_2(x) = c_1 \frac{\partial y}{\partial m} + c_2 \frac{\partial y}{\partial k} = c_1 x + c_2 \tag{11.9}$$

is the general solution to Jacobi's equation and that

$$\Delta(x, a) = u_2(a)\, u_1(x) - u_1(a)\, u_2(x) \tag{11.10}$$

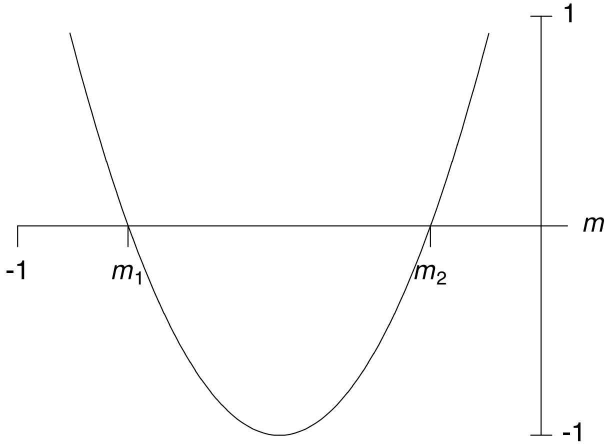



**Figure 11.1.** Plot of $R$

reduces to

$$\Delta(x, a) = (x - a). \tag{11.11}$$

This solution vanishes at $x = a$ but does not vanish again. There is no conjugate point and the strengthened Legendre condition is satisfied.

Our extremal satisfies the sufficiency conditions for a weak relative minimum for $m < m_1$ and for $m > m_2$. It satisfies the corresponding sufficiency conditions for a weak relative maximum for $m_1 < m < m_2$. Nevertheless, the extremal is not a minimum or a maximum (relative to strong variations) for $-1 < m < 0$.

The situation is most clearly seen with respect to minima. Consider the straight line

$$y = -\frac{1}{5}x + 2, \quad 0 \leq x \leq 5, \tag{11.12}$$

that is the extremal for the boundary conditions

$$y(0) = 2, \quad y(5) = 1. \tag{11.13}$$

In any, arbitrarily small, weak-norm, strong-variation neighborhood of this extremal we can do better by joining our endpoints with a broken line consisting of line segments of slope 0 and $-1$ (as suggested

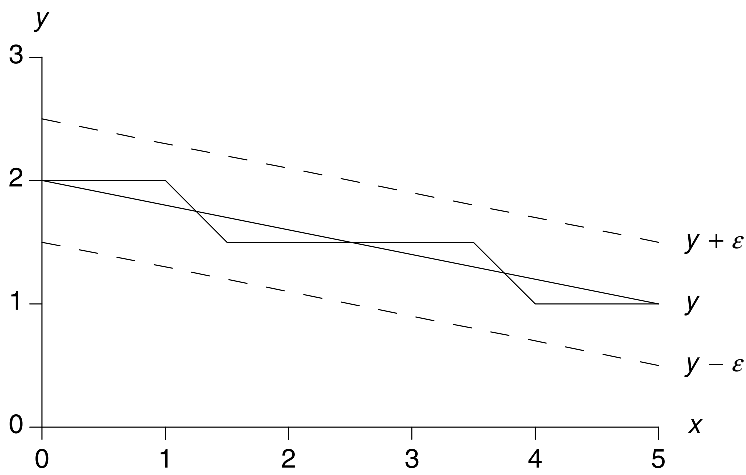



**Figure 11.2.** Broken variation about an extremal

by the characteristic or indicatrix). This broken curve will be mapped by the functional to zero, which is clearly the minimum for $J$.

You may argue that using a broken extremal is somehow unfair. However, we can always smooth the corners of this extremal to produce a continuously differentiable function whose integral is arbitrarily close to zero. The real problem is that the derivative of the difference between the broken extremal and the continuously differentiable straight-line extremal does not go to zero as $\epsilon$ goes to zero. There will always be parts of this difference with slope

$$-1 - \left(-\frac{1}{5}\right) \;=\; -\frac{4}{5}\,. \tag{11.14}$$

Don't think that the problem is tied exclusively to the existence of broken extremals. In effect, broken extremals are the tip of the iceberg. Once we allow broken extremals, it is only a small conceptual step to allow nearby comparison curves to have corners or large derivatives. That is, it is only a very small step to then consider strong variations. So, here is another, even more disturbing, example.

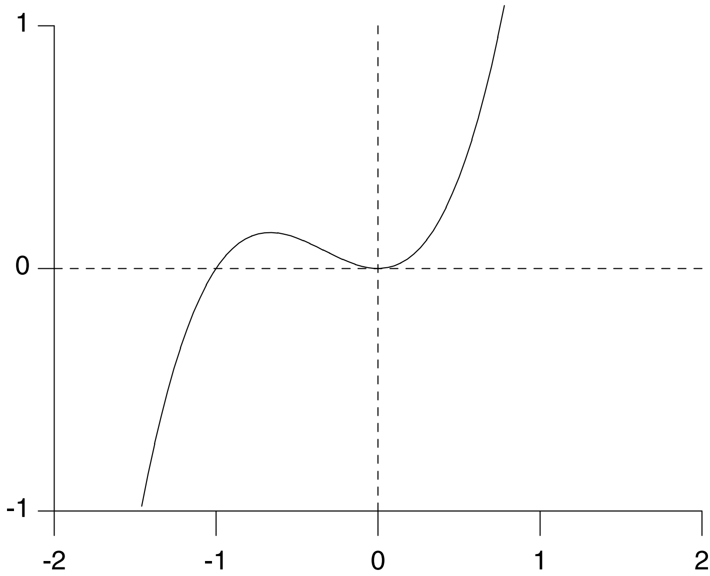



**Figure 11.3.** Indicatrix

**Example 11.2.**

Consider the functional

$$J[y] \;=\; \int_0^1 (y'^2 + y'^3)\; dx \qquad (11.15)$$

with boundary conditions

$$y(0) \;=\; 0\,, \quad y(1) \;=\; 0\,. \qquad (11.16)$$

The first integral for this problem is

$$\frac{\partial f}{\partial y'} \;=\; 2y' + 3y'^2 \;=\; \alpha\,. \qquad (11.17)$$

It follows that $y'$ is a constant and that the extremal is the straight line

$$y \;=\; 0 \;\;\text{for}\;\; 0 \,<\, x \,<\, 1\,. \qquad (11.18)$$

Along this extremal, $J[y]$ is equal to zero.

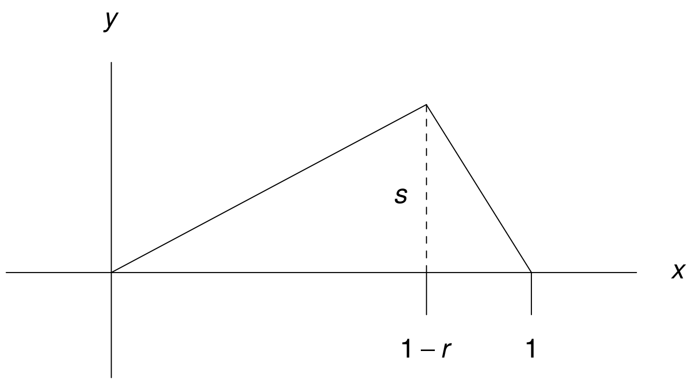



**Figure 11.4.** A jagged variation

For this extremal,

$$R = \left.\frac{\partial^2 f}{\partial y'^2}\right|_{y=0} = 2 \tag{11.19}$$

and so the strengthened Legendre condition for a minimum is satisfied. Also,

$$\Delta(x, 0) = x\,, \tag{11.20}$$

there is no conjugate point, and the strengthened Jacobi condition is satisfied. All of the conditions for weak relative minimum are satisfied. In addition, this problem has no broken extremals. A quick plot of the characteristic (see Figure 11.3) shows that there is no tangent line that is tangent to two distinct points of the characteristic.

In spite of all this, $J[y]$ can be made negative. Consider the function

$$y = \begin{cases} \dfrac{s\,x}{1-r}\,, & 0 \le x \le 1-r\,, \\[2ex] \dfrac{s\,(1-x)}{r}\,, & 1-r \le x \le 1 \end{cases} \tag{11.21}$$

(see Figure 11.4). For this function

$$J[y] = \int_0^{1-r} \frac{s^2}{(1-r)^2}\left(1 + \frac{s}{1-r}\right)\,dx + \int_{1-r}^{1} \frac{s^2}{r^2}\left(1 - \frac{s}{r}\right)\,dx \tag{11.22}$$

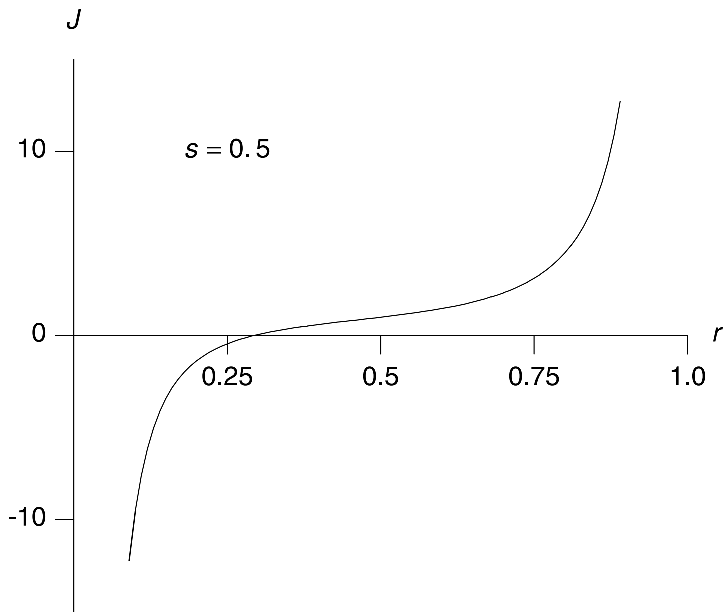



**Figure 11.5.** $J[y]$ for a sawtooth variation

so that

$$J[y] \quad = \quad \frac{s^2}{r(1-r)}\left[1+\frac{sr}{(1-r)}-\frac{s(1-r)}{r}\right]. \qquad (11.23)$$

The last term can be made more negative than other terms by making $r$ sufficiently small. Figure 11.5 shows a plot of this function for a typical value of $s$.

In this example, as we make the slope of the declining portion of our jagged or "sawtooth" variation more negative, we outperform our $x$-axis extremal. This occurs in spite of the fact that the $x$-axis extremal is a weak relative minimum. Our jagged sawtooth function is typical of the strong variations introduced by Weierstrass in 1879.

All of the earliest workers in the calculus of variations assumed that slopes of variations tend to zero as their ordinates vanish. In fact, there is no reason why the vanishing of the ordinate should imply the vanishing of the slope. In dealing with strong variations, we still keep all of our previous necessary conditions. At the same time, our old necessary conditions are not enough for strong variations. We need at least one new condition.

There is an additional condition. It is due to and named after Weierstrass. Weierstrass realized that he could account for strong variations using the transversality condition. (To the extent that the transversality condition arose in the context of the first variation, you could argue that Weierstrass still relied on weak variations. We will fix this in the next chapter.) Let us look at this new necessary condition.

## 11.2. Weierstrass's condition

Let us imagine that we have an extremal, $y = \hat{y}(x)$, without corners (see Figure 11.6). Let point 1, with coordinates $(x_1, y_1)$, be on this curve, and let point 3, with coordinates $(x_3, y_3)$, be on this curve to the right of point 1. Let an arbitrary admissible curve, $y = h(x)$, intersect the extremal at point 1. Let point 2, with abscissa $x_1 + \sigma$ be a point on $y = h(x)$, to the right of point 1. Note that the slope of curve $y = h(x)$ at point 1 — call it $q$ — differs, in general, by a finite amount from the slope of $\hat{y}(x)$ at point 1. This remains true even if we were to move point 2 closer to point 1 along $y = h(x)$. Let $y_{23}(x)$ be the extremal that connects points 2 and 3.

We will write $I_{12}$ to denote the value of our integral between points 1 and 2 along the arc $y_{12}$, $I_{23}$ to denote the value of our integral between points 2 and 3 along $y_{23}(x)$, and $I_{13}$ to denote the value of our integral between points 1 and 3 along $\hat{y}(x)$. By the variable-endpoint formula, we have

$$I_{23} - I_{13} \; = \; - \left[ \left( f - y' \frac{\partial f}{\partial y'} \right) + q \, \frac{\partial f}{\partial y'} \right] \sigma \, . \tag{11.24}$$

The values $x$, $y$, and $y'$ in this expression are evaluated along $\hat{y}$, at point 1. For small $\sigma$, we also have

$$I_{12} \; = \; f(x_1, \hat{y}_1, q) \, \sigma \, . \tag{11.25}$$

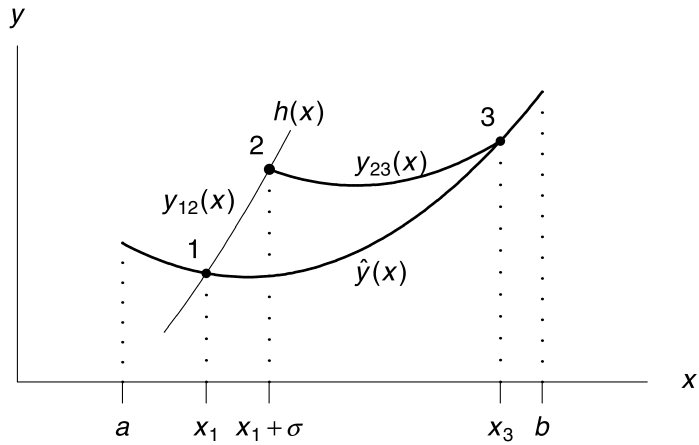



**Figure 11.6.** A strong variation

For the variation of the integral, we may thus write

$$I_{12} + I_{23} - I_{13} \qquad (11.26)$$
$$= \left[ f(x_1, \hat{y}_1, q) - f(x_1, \hat{y}_1, \hat{y}_1') - \frac{\partial f}{\partial y'}(x_1, \hat{y}_1, \hat{y}_1')(q - \hat{y}_1') \right] \sigma \, .$$

This expression must be nonnegative if $\hat{y}(x)$ is to be a minimum.

The point 1 was arbitrary in the above discussion. This being the case, let us consider

$$E(x, y, p, q) \;=\; f(x, y, q) - f(x, y, p) - \frac{\partial f}{\partial y'}(x, y, p)(q - p) \, . \quad (11.27)$$

This functions is known as the *Weierstrass E-function* or as the *Weierstrass excess function.* (The variable $p$ here is not momentum.) We may now write a new necessary condition in terms of this function.

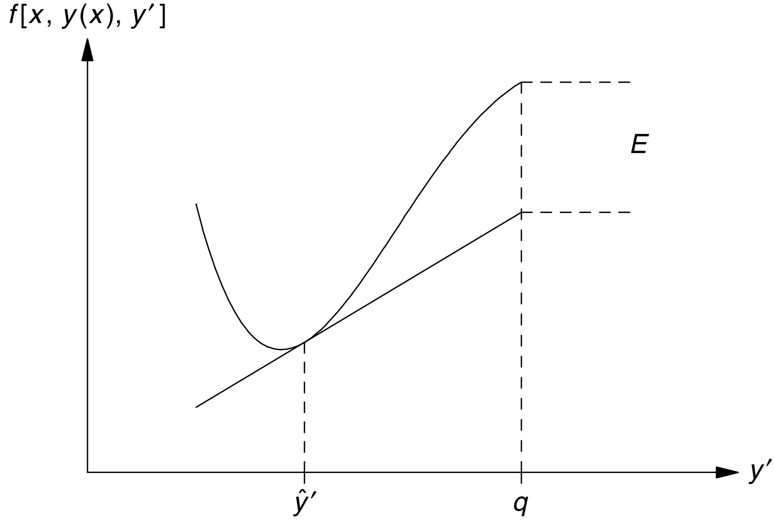



**Figure 11.7.** Weierstrass meets the indicatrix

> ***Weierstrass's condition:***
> For the functional $J[y]$ to have a strong relative minimum (maximum) at $y = \hat{y}(x)$ we must have that
> $$E(x, \hat{y}, \hat{y}', q) \ \geq \ 0 \ \ (\leq 0) \qquad (11.28)$$
> at every point of $y = \hat{y}(x)$ and for every finite value of $q$.

Weierstrass's condition has a simple geometric interpretation. Let us plot the characteristic or indicatrix for our problem (see Figure 11.7). The Weierstrass condition,

$$E(x, \hat{y}, \hat{y}', q) \ \geq \ 0\,, \qquad (11.29)$$

implies that the characteristic must lie above — or at least not below — the tangent line through $\hat{y}'$. This is a local convexity requirement on the integrand.

Here is a small table that lists all of the necessary conditions for a strong relative extremum that we have collected:

***Necessary conditions for a strong extremum:***

(a) Euler–Lagrange condition
(b) Legendre condition
(c) Jacobi condition
(d) Weierstrass condition

Let us now look at some simple examples of using Weierstrass's necessary condition.

**Example 11.3.**

We previously studied the functional

$$J[y] = \int_0^b f(x, y, y')\, dx = \int_0^b (y'^2 - y^2)\, dx \tag{11.30}$$

for $0 < b < \pi$ and for the boundary conditions

$$y(0) = 0\,, \quad y(b) = 1 \tag{11.31}$$

and concluded that the extremal

$$\hat{y}(x) = \frac{\sin x}{\sin b} \tag{11.32}$$

satisfies all the criteria to be a weak relative minimum. Let us now check Weierstrass's condition. For this problem, the excess function is

$$\begin{aligned} E(x, \hat{y}, \hat{y}', q) &= f(x, \hat{y}, q) - f(x, \hat{y}, \hat{y}') - \frac{\partial f}{\partial y'}(x, \hat{y}, \hat{y}')(q - \hat{y}') \qquad (11.33) \\ &= (q^2 - \hat{y}^2) - [(\hat{y}')^2 - \hat{y}^2] - 2\,\hat{y}'\,(q - \hat{y}') \\ &= q^2 - 2\,q\,\hat{y}' + (\hat{y}')^2 \\ &= (q - \hat{y}')^2\,. \end{aligned}$$

It is thus clear that

$$E(x, \hat{y}, \hat{y}', q) \geq 0 \tag{11.34}$$

and that Weierstrass's condition is satisfied.

**Example 11.4.**

For the functional

$$J[y] = \int_0^1 ({y'}^2 + {y'}^3)\, dx \tag{11.35}$$

with boundary conditions

$$y(0) = 0\,, \quad y(1) = 0\,, \tag{11.36}$$

we saw that the extremal of interest,

$$\hat{y}(x) = 0 \quad \text{for} \quad 0 < x < 1\,, \tag{11.37}$$

is a line of zero slope. The Weierstrass excess function for this extremal reduces to

$$E(x, \hat{y}, \hat{y}', q) = q^2(1 + q) \tag{11.38}$$

and it is clear that the excess function can be of either sign, depending on the sign and magnitude of $q$. Weierstrass's necessary condition is not satisfied and the extremal $y = 0$ is not a strong relative minimum.

**Example 11.5.**

For the functional

$$J[y] = \int_a^b (y' + 1)^2 {y'}^2\, dx\,, \tag{11.39}$$

the extremals are straight lines of slope $m$. The excess function for this set of extremals is just

$$\begin{aligned} E(x, \hat{y}, \hat{y}', q) &= q^2(q+1)^2 - m^2(m+1)^2 \\ &\quad - 2m(m+1)(2m+1)(q-m) \\ &= (q-m)^2\left[q^2 + 2(m+1)q + (3m^2 + 4m + 1)\right]. \end{aligned} \tag{11.40}$$

The sign of the excess function is controlled by the expression in square brackets. This expression is a quadratic in $q$ that has a minimum at

$$q = -(m + 1)\,. \tag{11.41}$$

(To the extent that we are dealing with a minimum, it should be immediately apparent that Weierstrass's condition for a maximum is not satisfied.) At this value of $q$, the expression in square brackets simplifies to

$$m(m+1)\,. \tag{11.42}$$

For $m < -1$ or $m > 0$, this minimum is positive and Weierstrass's necessary condition for a minimum is satisfied. For $-1 < m < 0$, this minimum is negative and Weierstrass's condition for a minimum (relative to strong variations) is not satisfied.

## 11.3. Case study: Newton's problem

Weierstrass's use of strong variations and the excess function did not occur in a vacuum. Rather, they occurred as part of Weierstrass's effort to understand a classic problem that had plagued the calculus of variations in the second half of the 19th century. That classic conundrum was Newton's problem of finding the solid of revolution that moves through an inviscid and incompressible fluid with least resistance.

Newton formulated his problem in 1685. Newton then published this problem, and his solution, in his famous *Principia Mathematica* of 1687 — nine years before John Bernoulli's brachistochrone challenge of 1696. So, Newton's problem has a legitimate claim to being the first genuine problem of the calculus of variations. Nevertheless, Newton's problem did not have the same immediate impact that the brachistochrone had.

Newton published his solution with no hint as to his method of solution and it is fair to say that Newton's solution mystified the European mathematical community (Goldstine, 1980). Finally, David Gregory persuaded Newton to write out his method of solution; Gregory then presented lectures on this material at Oxford. Let us now look at Newton's problem in detail to see why it proved so troublesome to mathematicians.

Let us assume that a solid of revolution moves, with constant velocity $v$, through a perfectly inviscid and incompressible fluid (see

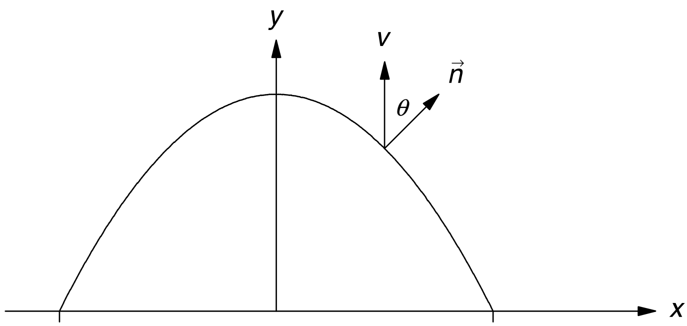



**Figure 11.8.** Newton's problem

Figure 11.8). The solid moves in the direction of its axis of rotation, which we will take, for convenience, as the $y$-axis. There is, by assumption, no friction between the solid body and the fluid, but we follow Newton in assuming that the resistance on any element of surface is proportional to the square of the normal component of the velocity.

Since we are dealing with a surface of revolution, a small ring of surface has the area

$$2\pi x\ ds \;=\; 2\pi x\sqrt{1+y'^2}\ dx\,, \tag{11.43}$$

where $x$ is the radius and $ds$ is the element of arc length for our solution curve, $y = y(x)$, for that ring. The component of velocity in the normal direction to the surface is $v\cos\theta$ and the resistance in the normal direction offered by the zone is therefore

$$v^2\cos^2\theta\cdot 2\pi x\sqrt{1+y'^2}\ dx\,. \tag{11.44}$$

This quantity is multiplied by $\cos\theta$ once again, to give the portion of the resistance in the $y$ direction, since the $x$ component of the resistance serves merely to pinch the solid. Thus, the relevant resistance is proportional to the integral of

$$2\pi v^2 x\cos^3\theta\,\sqrt{1+y'^2}\ dx\,. \tag{11.45}$$

It is easy to show, moreover, that

$$\cos\theta \;=\; \frac{1}{\sqrt{1+y'^2}} \tag{11.46}$$

so that the total resistance is proportional to

$$J[y] \;=\; \int \frac{x}{1+y'^2}\,dx\,. \tag{11.47}$$

I have left the limits of integration and the boundary conditions unspecified, for reasons that will become clear shortly.

The above integrand does not have any explicit $y$ dependence and so we can immediately write down the first integral

$$\frac{\partial f}{\partial y'} \;=\; -\frac{2xy'}{(1+y'^2)^2} \;=\; 2\,c \tag{11.48}$$

with the constant of integration $2c$. As a result,

$$-\frac{x\,y'}{(1+y'^2)^2} \;=\; c\,. \tag{11.49}$$

In general, for nonzero $c$, we must now solve a quartic equation if we wish to solve for $y'$. Rather than taking that direct approach, we will instead determine a parametric solution. Figure 11.8 shows the putative solution curve decreasing in $y$ and increasing in $x$ for $x > 0$. For consistency, we now let

$$p \;\equiv\; -\frac{dy}{dx} \tag{11.50}$$

so that $c$ and $p$ are positive (in Figure 11.8) for $x > 0$. Our first integral, equation (11.48), now implies that

$$x \;=\; c\,\frac{(1+p^2)^2}{p}\,. \tag{11.51}$$

Moreover, since

$$dy \;=\; -p\;dx\,, \tag{11.52}$$

it quickly follows that

$$dy \;=\; c\left(\frac{1}{p} - 2p - 3p^3\right) \tag{11.53}$$

and that

$$y = y_0 + c\left(\ln p - p^2 - \frac{3}{4}p^4\right). \tag{11.54}$$

Equations (11.51) and (11.54) define a possible solution curve.

What does the trace of this curve look like? Notice that

$$\frac{dx}{dp} = c\,\frac{(1+p^2)(3p^2-1)}{p^2}, \tag{11.55}$$

$$\frac{dy}{dp} = -c\,\frac{(1+p^2)(3p^2-1)}{p} \tag{11.56}$$

and that both of these derivatives vanish at

$$p = \frac{\sqrt{3}}{3}. \tag{11.57}$$

We may thus expect a cusp (or some other, higher-order singularity) at this value of $p$, at the coordinates

$$x = \frac{16\sqrt{3}}{9}c, \quad y = y_0 - \left(\frac{1}{2}\ln 3 + \frac{5}{12}\right)c. \tag{11.58}$$

Figure 11.9 shows our solution curve plotted for $c = 0.75$ and $y_0 = 5$. We do indeed see two branches coming together at a cusp. The upper branch results from small values of $p$: this branch comes in from $x = \infty$ (at $p = 0$) into the cusp (at $p = \sqrt{3}/3$). The lower branch corresponds to higher values of $p$: the lower branch starts at the cusp, for $p = \sqrt{3}/3$, and goes off to $y = -\infty$, as $p$ increases.

If we look at the parametric and geometric form of the solution curve, we also see that some boundary conditions are incompatible with this curve and with this problem. Based on Figure 11.8, you might have guessed that we would have boundary points on each axis,

$$y(0) = h, \quad y(w) = 0, \tag{11.59}$$

so that we specify the height $h$ and width (or radius) $w$ of the solid. Unfortunately, imposing the boundary condition $y(0) = h$ would then force $c = 0$ and $y = y_0 = h$, which would not, in general, satisfy the second boundary condition. Indeed, for the solution of Newton's problem to make sense, we must modify the problem, as Newton did, so that the first boundary point is off the $y$-axis (see Figure 11.10). We must, in other words, turn the problem into one of determining

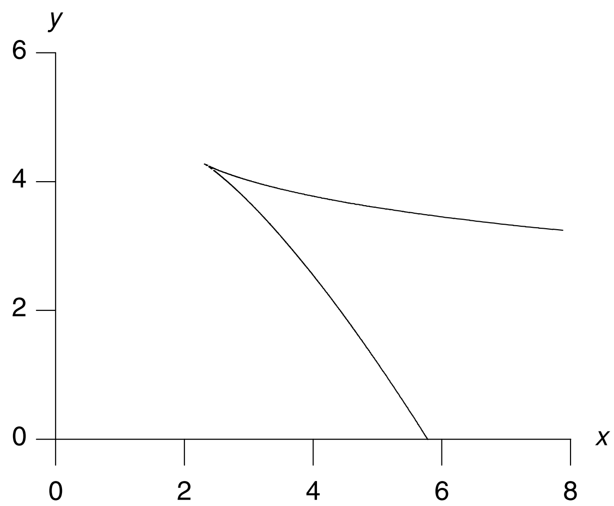



**Figure 11.9.** Two branches and a cusp

the profile of the *shoulder* of the solid of revolution that minimizes resistance, with the very tip of the projectile prespecified. Newton took this tip to be flat. Our new boundary conditions are

$$y(a) = h\,, \quad y(w) = 0, \tag{11.60}$$

with $a > 0$.

We now have boundary conditions that our solution curve can satisfy. At the same time, our solution curve has two branches. We need to verify which, if either, of the two branches corresponds to a minimum. Let us start with the Legendre condition. Since

$$\frac{\partial^2 f}{\partial y'^2} = \frac{2x(3y'^2 - 1)}{(1 + y'^2)^3} = \frac{2x(3p^2 - 1)}{(1 + p^2)^3}\,, \tag{11.61}$$

we see that the lower branch, corresponding to $p \geq \sqrt{3}/3$, is consistent with a weak relative minimum, but that the upper branch, corresponding to $0 < p \leq \sqrt{3}/3$, is consistent with a weak relative

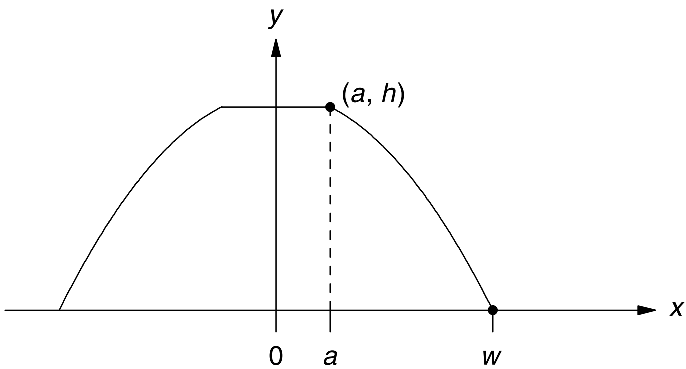



**Figure 11.10.** The revised problem

maximum. Verifying the Jacobi condition is more difficult, but with due diligence one can, in fact, show that the Jacobi test imposes no limit on the range of a solution that starts on the lower branch of our cusped solution curve.

We have a solution, Newton's solution, that satisfies all the necessary conditions and, more often than not, all the sufficient conditions for a weak relative minimum. And yet, it has been known, since the 18th century, that one can construct solids of revolution that have less resistance than the solid suggested by Newton. The first to call attention to this fact appears to have been Guillaume de Saint-Jacques de Silvabelle, the director of the observatory in Marseilles. Legendre noted this in his famous paper on the Legendre condition (Legendre, 1788). In 1760, Saint-Jacques de Silvabelle claimed that he could draw a polygonal line (see Figure 11.11) that generated less resistance than the solution of the Euler–Lagrange equation. If we reexamine our original integral, equation (11.47), Saint-Jacques de Silvabelle's claim makes good sense: by making $y'(x)$ large enough, we can make our integrand and integral arbitrarily small.

By the middle of the 19th century, Newton's problem was viewed as paradoxical. Mathematicians responded by expelling this problem

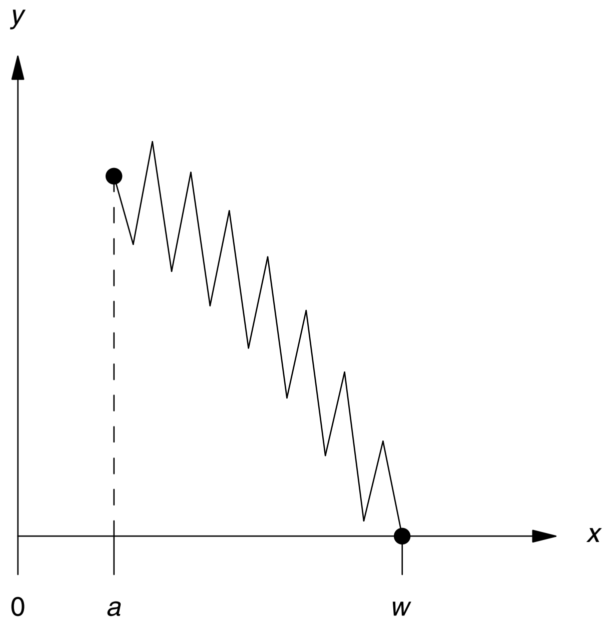



**Figure 11.11.** A polygonal profile

from their textbooks (Kolmogorov and Yushkevich, 1998). In hindsight, it is easy to see that Saint-Jacques de Silvabelle's solution consists of strong variations and that the conditions of Euler, Legendre, and Jacobi do not necessarily guarantee a minimum relative to strong variations. Indeed, if we look at the Weierstrass excess function for this problem,

$$\begin{aligned} E(x,\hat{y},\hat{y}',q) &= f(x,\hat{y},q) - f(x,\hat{y},\hat{y}') - \frac{\partial f}{\partial y'}(x,\hat{y},\hat{y}')(q-\hat{y}') \\ &= \frac{x\,(q-\hat{y}')^2\,[2\hat{y}'q + (\hat{y}')^2 - 1]}{[1+(\hat{y}')^2]^2\,(1+q^2)}, \end{aligned} \tag{11.62}$$

we see that we cannot force the excess function to be nonnegative for all choices of $q$. As a result, the integral possesses a weak minimum, but not a strong minimum. Weierstrass needed to expand the theory of the calculus of variations to account for the difficulties associated with Newton's problem. He did so, in his lectures, by introducing strong variations and his excess function.

## 11.4. Recommended reading

See Graves (1934) for an alternative derivation of Weierstrass's necessary condition.

Kolmogorov and Yushkevich (1998) discuss the importance of Newton's problem to the calculus of variations. Newton's problem continues to fascinate scientists. Recent papers that consider this problem include Buttazzo and Kawohl (1993), Horstmann et al. (2002), Lachand-Robert and Oudet (2005), Silva and Torres (2006), and Cruz-Sampedro and Tetlalmatzi-Montiel (2014). There are many others. See Miele (1965) for related problems in aerodynamics.

## 11.5. Exercises

**11.5.1. Checking necessary conditions.** Consider the integral

$$J[y] \;=\; \int_a^b y^2(1-y'^2)\;dx\,. \tag{11.63}$$

Show that the Legendre and Weierstrass necessary conditions for a local maximum are satisfied by all extremals.

**11.5.2. A cubic integrand.** Consider the integral

$$J[y] \;=\; \int_0^1 y'^3\,dx \tag{11.64}$$

with the boundary conditions

$$y(0) \;=\; 0\,,\quad y(1) \;=\; 1\,. \tag{11.65}$$

Show that $\hat{y}(x) = x$ is an extremal that satisfies the Legendre condition but not the Weierstrass condition for a local minimum.

# Chapter 12

# Sufficient Conditions

## 12.1. Introduction

The necessary conditions of Euler, Legendre, and Jacobi were published in 1744, 1788, and 1837. These conditions are not, by themselves, sufficient conditions for a strong relative minimum. In 1879, Weierstrass added a fourth necessary condition, built upon his excess function. Weierstrass also realized that he could strengthen and combine the four necessary conditions to derive a package of sufficient conditions that do guarantee the presence of a strong relative minimum.

In 1900, David Hilbert simplified the proof of Weierstrass's sufficient conditions. Our goal, in this chapter, is to understand Weierstrass's sufficient conditions and Hilbert's proof of these sufficient conditions. We will then look at Carathéodory's method of equivalent variational problems. Carathéodory's elegant method is sometimes called the royal road to the calculus of variations. To follow Hilbert's and Carathéodory's approaches, we must first understand what is meant by a "field of extremals."

## 12.2. Fields of extremals

Let $D$ be a domain in the $(x, y)$ plane and let

$$y = \phi(x, c) \tag{12.1}$$

be a one-parameter family of curves that covers $D$. If one, and only one, member of this family passes through each point of $D$, then this family of curves is called a *proper field* on the domain $D$. The fact that a field is proper implies that there exists a single-valued function,

$$c = \psi(x, y)\,, \tag{12.2}$$

such that

$$y = \phi(x, \psi(x, y)) \tag{12.3}$$

for every point in our domain.

For a proper field, the slope of the tangent to the curve $y = \phi(x, c)$ at each point now defines a function, $p(x, y)$, that we call the *slope of the field.* This slope is defined, analytically, by the two equations

$$p(x, y) = \frac{\partial \phi}{\partial x}(x, c)\,, \quad c = \psi(x, y)\,. \tag{12.4}$$

**Example 12.1.**

Consider the unit disk

$$x^2 + y^2 \leq 1\,. \tag{12.5}$$

The set of parallel straight lines

$$y = x + c \tag{12.6}$$

is a proper field on the disk with slope $p(x, y) = 1$ (see Figure 12.1). The function $\psi(x, y)$ is simply

$$c = \psi(x, y) = y - x\,. \tag{12.7}$$

The one-parameter family of parabolas

$$y = (x - c)^2 - 1 \tag{12.8}$$

is not, in contrast, a proper field on the disk since two parabolas may, in general, pass through the same point.

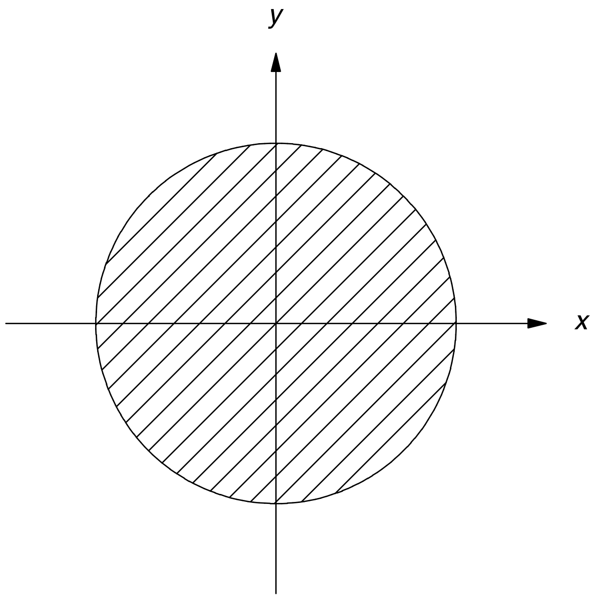



**Figure 12.1.** A proper field

If all the curves of a one-parameter family of curves pass through the same point and form a pencil of curves, they do not, of course, form a proper field. If, however, these curves cover the whole domain $D$ and never intersect each other except at the center or nib of the pencil, the curves are said to constitute a *central field.*

**Example 12.2.**

The pencil of sine curves

$$y = \phi(x, c) = c \sin x \tag{12.9}$$

(see Figure 12.2) is a central field for sufficiently small neighborhoods of the strip $0 \le x \le b$, for $b < \pi$. It is a proper field for the strip $a \le x \le b$, for $a > 0$ and $b < \pi$.

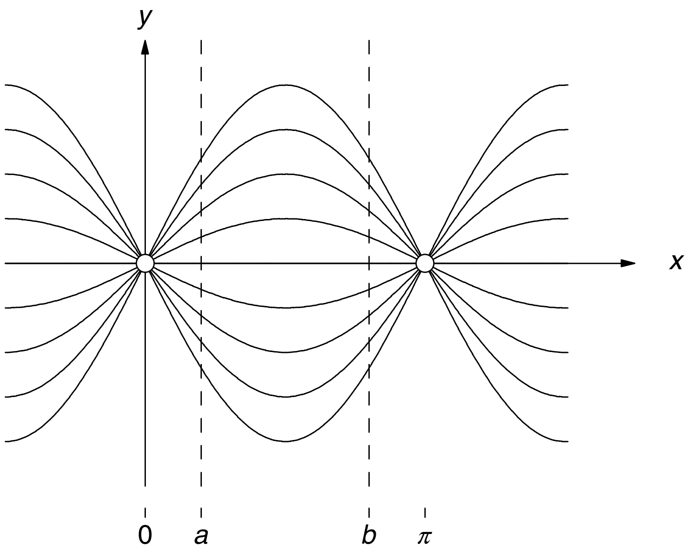



**Figure 12.2.** A pencil of sine curves

If we have a proper or central field that is also a one-parameter family of extremals for a variational problem, we have a *field of extremals.*

Our ability to cover a domain with a field of extremals is closely related to the Jacobi condition. For simple problems, the easiest way to generate a one-parameter family of extremals is to start with the general solution of the Euler–Lagrange equation and to impose one of the boundary conditions, say at point $A$. This then gives us a one-parameter family of extremals, $y = \phi(x, c)$, emanating from point $A$. One of these extremals may also satisfy the second boundary condition at, say, point $B$. Let us call this extremal $\hat{y}(x)$. We have seen, in Chapter 6, that two infinitesimally close members of a one-parameter family emanating from $A$ will intersect at a conjugate point lying on the envelope of the family of curves (see Figure 12.3). The conjugate point lies on the c-discriminant defined by the two equations

$$y = \phi(x, c)\,, \quad \frac{\partial y}{\partial c}(x, c) = 0\,. \tag{12.10}$$

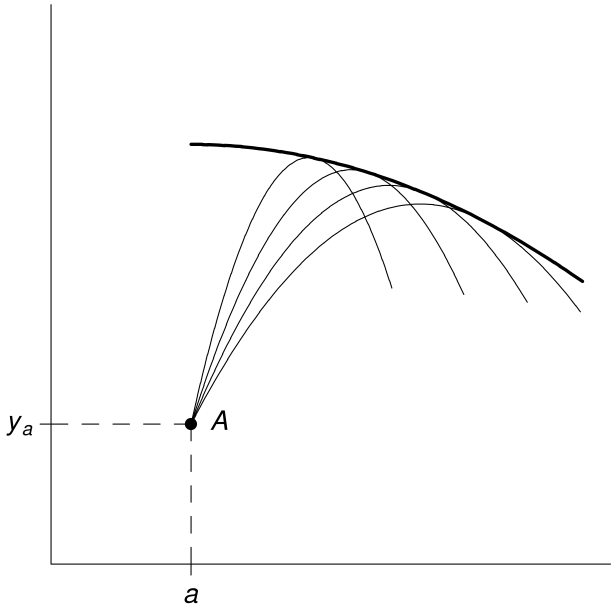



**Figure 12.3.** An envelope

If extremal $\hat{y}(x)$ does not touch this envelope, nearby extremals (sufficiently close members of our one-parameter family) do not intersect $\hat{y}(x)$. We then have a central field that includes $\hat{y}(x)$ and that covers some neighborhood of $\hat{y}(x)$. If $\hat{y}(x)$ does touch the envelope, nearby extremals will intersect $\hat{y}(x)$ and we do not have a field. We must thus satisfy the Jacobi condition, actually the strengthened Jacobi condition, in order to have a field of extremals around a prescribed extremal.

For a field of extremals, we can say more about the slope of the field, $p(x, y)$. Differentiating slope equations (12.4), we see that

$$p_x = \phi_{xx} + \phi_{xc}\, c_x\,, \quad p_y = \phi_{xc}\, c_y\,. \tag{12.11}$$

Straightforward implicit differentiation of field equation (12.1), in turn, implies that

$$c_x = -\frac{\phi_x}{\phi_c}\,, \quad c_y = \frac{1}{\phi_c}\,. \tag{12.12}$$

From our slope equations (12.4) and derivatives (12.11) and (12.12), it now follows that

$$p_x + p\,p_y \;=\; \phi_{xx}\,. \tag{12.13}$$

Since $\phi(x, c)$, however, is an extremal and satisfies the Euler–Lagrange equation for every value of $c$, we know, from the ultradifferentiated form of Chapter 3, that

$$f_{y'y'}\phi_{xx} + f_{y'y}\phi_x + f_{y'x} - f_y \;=\; 0\,. \tag{12.14}$$

Here, the arguments of the partial derivatives of the integrand $f$ of our functional are $x$, $\phi(x, c)$, and $\phi_x(x, c)$. It now follows that the slope $p(x, y)$ must satisfy the first-order partial differential equation

$$(p_x + pp_y)f_{y'y'} + pf_{y'y} + f_{y'x} - f_y \;=\; 0\,. \tag{12.15}$$

We will see this same partial differential equation arise, in a very different way, in the next section.

## 12.3. Hilbert's invariant integral

Let us return to the problem of finding the extremum of the functional

$$J[y] \;=\; \int_a^b f(x, y(x), y'(x))\;dx \tag{12.16}$$

subject to the boundary conditions

$$y(a) \;=\; y_a\,, \quad y(b) \;=\; y_b\,. \tag{12.17}$$

We will assume that we have found an extremal, $y = \hat{y}(x)$, that satisfies both boundary conditions and that this extremal is surrounded by a field of extremals. This implies that $y = \hat{y}(x)$ satisfies both the Euler–Lagrange equation and the strengthened Jacobi condition.

Let us now examine the total variation

$$\Delta J \;=\; J[y] - J[\hat{y}] \tag{12.18}$$

or

$$\Delta J \;=\; \int_a^b f(x, y(x), y'(x))\;dx - \int_a^b f(x, \hat{y}(x), \hat{y}'(x))\;dx \tag{12.19}$$

for curves, $y = y(x)$, that satisfy our boundary conditions and that lie in the domain covered by our field. We require $\Delta J \geq 0$ for a minimum and $\Delta J \leq 0$ for a maximum. In Chapter 2, we wrote $y(x) = \hat{y}(x) + \epsilon\,\eta(x)$ and expanded the total variation in a power series in $\epsilon$. We are no longer free to follow this approach since we are now allowing strong variations and have no guarantee that $\eta'(x)$ is small.

We will instead follow Hilbert (1902) and replace the second integral in equation (12.19) by an equivalent, path-independent integral,

$$\int_a^b \Phi(x, y(x), y'(x))\ dx\,. \tag{12.20}$$

We want this integral to assume the value $J[\hat{y}(x)]$, not just for $y = \hat{y}(x)$, but for all $y = y(x)$ that satisfy our boundary conditions and that lie in the domain covered by our field. What form should $\Phi(x, y, y')$ take?

Hilbert took $\Phi(x, y, y')$ to be of the general form

$$\Phi(x, y, y') \;=\; f(x, y, p) + (y' - p) f_{y'}(x, y, p)\,, \tag{12.21}$$

with $p = p(x, y)$ an, as yet, undetermined function of $x$ and $y$. Note that $\Phi(x, y, y')$ is of the general form

$$\Phi(x, y, y') \;=\; M(x, y) + N(x, y)\, y'\,, \tag{12.22}$$

with, in this instance,

$$M(x, y) \;=\; f(x, y, p) - p\, f_{y'}\,(x, y, p)\,, \tag{12.23}$$

$$N(x, y) \;=\; f_{y'}(x, y, p)\,. \tag{12.24}$$

We saw, in Chapter 3, that integrands of this form yield path-independent integrals if

$$\frac{\partial M}{\partial y} \;=\; \frac{\partial N}{\partial x}\,. \tag{12.25}$$

Since

$$\frac{\partial M}{\partial y} \;=\; f_y - p\,(f_{y'y} + f_{y'y'} p_y) \tag{12.26}$$

and

$$\frac{\partial N}{\partial x} \;=\; f_{y'x} + f_{y'y'} p_x \,, \tag{12.27}$$

we now demand that $p = p(x, y)$ satisfy

$$(p_x + p\, p_y) f_{y'y'} + p f_{y'y} + f_{y'x} - f_y \;=\; 0\,. \tag{12.28}$$

This equation is identical, however, to the partial differential equation for the slope of our field of extremals, equation (12.15). So, if we choose $p = p(x, y)$ to be this slope, the integral

$$\int_a^b [f(x, y, p) + (y' - p) f_{y'}(x, y, p)]\; dx \tag{12.29}$$

is path-independent. We let $p = p(x, y)$ be the slope of our field of extremals in all that follows.

For $y = \hat{y}(x)$ and $y' = \hat{y}'(x)$, we now have $p(x, y) = \hat{y}'(x)$ (along $y = \hat{y}(x)$) and integral (12.29) reduces to $J[\hat{y}]$. Since integral (12.29) is path-independent, it also has the value $J[\hat{y}]$ for all curves, $y = y(x)$, that satisfy our boundary conditions and that lie in the domain covered by our field.

## 12.4. Weierstrass's $E$-function revisited

If we now replace the second integral in total variation (12.19) with Hilbert's invariant integral, equation (12.29), we find that

$$\Delta J \;=\; \int_a^b [f(x, y, y') - f(x, y, p) - (y' - p) f_{y'}(x, y, p)]\; dx\,, \tag{12.30}$$

where the integration is now over an arbitrary admissible curve that lies in the domain covered by our field of extremals. Our integrand, however, is none other than the Weierstrass excess function,

$$E(x, y, p, y') \;=\; f(x, y, y') - f(x, y, p) - (y' - p) f_{y'}(x, y, p)\,, \tag{12.31}$$

and so we may now write

$$\Delta J = \int_a^b E(x, y, p, y')\, dx\,. \tag{12.32}$$

We now require that the excess function be nonnegative as part of a sufficient condition for a minimum. For, if $E(x, y, p, y') \geq 0$, then $\Delta J \geq 0$. For a maximum, we instead require that the excess function be nonpositive since, if $E(x, y, p, y') \leq 0$, then $\Delta J \leq 0$.

For a weak extremum, one or the other of these conditions must be satisfied for all $x$ and $y$ that are close to the extremal $\hat{y}(x)$ and for all values of $y'(x)$ that are close to $p(x, y) = \hat{y}'(x)$. Adding these new conditions to our earlier requirements provides us with sufficient conditions for weak extrema.

***Sufficient conditions for a weak extremum:***

(1) $\hat{y}(x)$ is an extremal, i.e., a solution of the Euler–Lagrange equation, satisfying the prescribed boundary conditions.
(2) The extremal $\hat{y}(x)$ can be embedded in a field of extremals. (This condition can be replaced by the strengthened Jacobi condition.)
(3) The Weierstrass excess function, $E(x, y, p, y')$, is of constant sign for all $(x, y)$ sufficiently close to the extremal $\hat{y}(x)$ and for all values of $y'(x)$ sufficiently close to $p(x, y) = \hat{y}'(x)$. For a minimum, we need $E(x, y, p, y') \geq 0$; for a maximum, we need $E(x, y, p, y') \leq 0$.

For a strong extremum, we once again require $E \geq 0$ (for a minimum) or $E \leq 0$ (for a maximum) for all values of $x$ and $y$ that are close to the extremal $\hat{y}(x)$, but now for all $y'(x)$, not just for those close to $p(x, y) = \hat{y}'(x)$. Adding these new conditions to our earlier requirements provides us with sufficient conditions for strong extrema.

> ***Sufficient conditions for a strong extremum:***
>
> (1) $\hat{y}(x)$ is an extremal, i.e., a solution of the Euler–Lagrange equation, satisfying the prescribed boundary conditions.
> (2) The extremal $\hat{y}(x)$ can be embedded in a field of extremals. (This condition can be replaced by the strengthened Jacobi condition.)
> (3) The Weierstrass excess function, $E(x, y, p, y')$, is of constant sign for all $(x, y)$ sufficiently close to the extremal $\hat{y}(x)$ and for all values of $y'(x)$. For a minimum, we need $E(x, y, p, y') \geq 0$; for a maximum, we need $E(x, y, p, y') \leq 0$.

**Example 12.3.**

Consider the functional

$$J[y] \;=\; \int_0^1 y'^3 \, dx \tag{12.33}$$

with the boundary conditions

$$y(0) \;=\; 0\,, \quad y(1) \;=\; 1\,. \tag{12.34}$$

Since the dependent variable is missing, the Euler–Lagrange equation for this problem reduces to

$$\frac{\partial f}{\partial y'} \;=\; 3y'^2 \;=\; \alpha\,. \tag{12.35}$$

It follows that $y'$ must be a constant. The extremals for this problem are thus straight lines of the form

$$y \;=\; mx + k\,. \tag{12.36}$$

We are especially interested in the extremal,

$$\hat{y}(x) \;=\; x\,, \tag{12.37}$$

that connects the two boundary points. The pencil of lines

$$y \;=\; mx \tag{12.38}$$

with center $(0, 0)$ is a central field that includes $\hat{y}(x) = x$. The slope field for this field is $p(x, y) = m$. Along $\hat{y}(x) = x$, $p(x, y) = 1$.

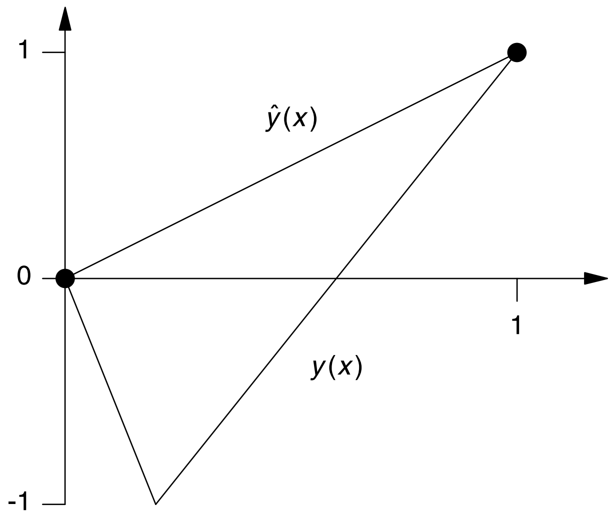



**Figure 12.4.** A broken line

The Weierstrass excess function for this problem is

$$\begin{aligned} E(x,y,p,y') &= f(x,y,y') - f(x,y,p) - (y'-p)f_{y'}(x,y,p) \\ &= {y'}^3 - m^3 - 3m^2(y'-m) \\ &= (y'-m)^2(y'+2m)\,. \end{aligned} \tag{12.39}$$

For $y'$ close to $m = 1$, $E \geq 0$ and all of the sufficient conditions for a weak relative minimum are satisfied. Therefore, $\hat{y}(x) = x$ is a weak relative minimum. At the same time, $\hat{y}(x) = x$ is not a strong relative minimum: if $y'$ is arbitrary, the sign of the excess function will not remain constant.

Indeed, let us compare the value of the functional along the extremal $\hat{y}(x) = x$ with the value of our functional along the broken curve

$$y = \begin{cases} -5x\,, & 0 \leq x \leq 0.2\,, \\ 2.5x - 1.5\,, & 0.2 \leq x \leq 1 \end{cases} \tag{12.40}$$

(see Figure 12.4). Along $\hat{y} = x$,

$$J[\hat{y}(x)] = \int_0^1 1\ dx = 1\,. \tag{12.41}$$

Along the broken curve,

$$J[y(x)] = \int_0^{0.2} (-5)^3\,dx + \int_{0.2}^1 (2.5)^3\,dx = -12.5\,. \tag{12.42}$$

Clearly, $J[y(x)] < J[\hat{y}(x)]$.

Checking sufficiency using the Weierstrass excess function can be cumbersome. It is useful, therefore, to have a simpler condition. We saw, in Chapter 11, that the Weierstrass condition is, in effect, a local convexity condition on the integrand of our variational problem. For purposes of proving sufficiency, we can replace local convexity with a stronger condition, global convexity. Here, for example, is a nice set of sufficient conditions for a strong minimum.

***Sufficient conditions for a strong minimum:***

(1) $\hat{y}(x)$ is an extremal, i.e., a solution of the Euler–Lagrange equation, satisfying the prescribed boundary conditions.

(2) The extremal $\hat{y}(x)$ contains no conjugate points.

(3) At all points on and in some neighborhood of the extremal and for all finite values of $y'$,

$$\frac{\partial^2 f}{\partial y'^2}(x, y, y') > 0\,. \tag{12.43}$$

Note the distinction between condition (12.43) and the strengthened Legendre condition. Inequality (12.43) must apply for ***all*** $y'$, not just those along the actual extremal.

**Example 12.4.**

For Fermat-type integrals with integrands of the form

$$f = g(x, y)\sqrt{1 + y'^2}\,, \tag{12.44}$$

it is easy to show that

$$\frac{\partial^2 f}{\partial y'^2} = \frac{g(x,y)}{(\sqrt{1+y'^2})^3} . \qquad (12.45)$$

Hence every extremal which does not contain a conjugate point provides a strong relative minimum provided that

$$g(x,y) > 0 \qquad (12.46)$$

along the extremal. This general class of integrals includes distance, the brachistochrone, and the minimal surface of revolution.

## 12.5. The royal road

Carathéodory (1935) introduced the method of equivalent variational problems. This method is one of the quickest and most elegant ways of deriving sufficient conditions for the calculus of variations, so much so that Carathéodory's method has been called the royal road to the calculus of variations (Boerner, 1953). Carathéodory's method also highlights the connection between the calculus of variations and classical Hamilton–Jacobi theory. Let us traverse the royal road.

Let us start with the problem of minimizing the functional

$$J[y] = \int_a^b f(x, y(x), y'(x)) \, dx \qquad (12.47)$$

subject to the boundary conditions

$$y(a) = y_a , \quad y(b) = y_b . \qquad (12.48)$$

Let $S(x,y)$ be any twice continuously differentiable function and let $C$ be any piecewise continuously differentiable contour, $y = y(x)$, that satisfies boundary conditions (12.48). Along the contour $C$,

$$\begin{aligned} \int_a^b \left( \frac{\partial S}{\partial x} + \frac{\partial S}{\partial y} y' \right) dx &= \int_a^b \frac{dS}{dx} \, dx \qquad (12.49) \\ &= S(b, y_b) - S(a, y_a) . \end{aligned}$$

Let us next introduce

$$f^*(x,y,y') \equiv f(x,y,y') - \frac{\partial S}{\partial x} - \frac{\partial S}{\partial y} y' . \qquad (12.50)$$

Along $C$, the functional

$$J^*[y] = \int_a^b f^*(x, y, y')\, dx \tag{12.51}$$

$$= \int_a^b f(x, y, y')\, dx - [S(b, y_b) - S(a, y_a)]$$

differs from our original functional, $J[y]$, by a mere constant. Any contour that minimizes $J[y]$ thus also minimizes $J^*[y]$. The two variational problems, with integrands $f(x, y, y')$ and $f^*(x, y, y')$, are said to be *equivalent.*

Every possible $S(x, y)$ generates an equivalent variational problem. Carathéodory sought an $S(x, y)$ and the slope of a field $p(x, y)$ that make the new variational problem especially easy to solve. In particular, he demanded that $S(x, y)$ and $p(x, y)$ be chosen so that

$$f^*(x, y, y') = 0, \text{ for } y' = p(x, y), \tag{12.52}$$

$$f^*(x, y, y') \geq 0, \text{ for } y' \neq p(x, y), \tag{12.53}$$

or, equivalently, so that

$$f(x, y, y') - \frac{\partial S}{\partial x} - \frac{\partial S}{\partial y} y' = 0, \text{ for } y' = p(x, y), \tag{12.54}$$

$$f(x, y, y') - \frac{\partial S}{\partial x} - \frac{\partial S}{\partial y} y' \geq 0, \text{ for } y' \neq p(x, y), \tag{12.55}$$

for all $x$ and $y$ (or at least all $x$ and $y$ sufficiently close to the solution).

If we want to minimize functional (12.51), it is now enough to choose, as our solution, the curve that satisfies

$$y' = p(x, y) \tag{12.56}$$

subject to the boundary conditions

$$y(a) = y_a, \quad y(b) = y_b. \tag{12.57}$$

Along this minimizing curve,

$$\int_a^b f(x, y, y')\, dx = S(b, y_b) - S(a, y_a). \tag{12.58}$$

Since the functional $f^*(x, y, y')$ assumes a minimum for $y' = p(x, y)$, the derivative of the left-hand side of equation (12.54) with respect to $y'$ must vanish for $y' = p(x, y)$. It now follows that

$$\frac{\partial S}{\partial y} = \frac{\partial f}{\partial y'}(x, y, p(x, y)) . \tag{12.59}$$

Equation (12.54) may now, in turn, be written as

$$\frac{\partial S}{\partial x} = f(x, y, p(x, y)) - p(x, y) \frac{\partial f}{\partial y'}(x, y, p(x, y)) . \tag{12.60}$$

Carathéodory thought these last two equations so important that he called them the *fundamental equations of the calculus of variations.* Others call them the *Carathéodory equations.*

If the Carathéodory equations, equations (12.59) and (12.60), are used to eliminate $S_x$ and $S_y$ in inequality (12.55), this inequality simplifies to

$$f(x, y, y') - f(x, y, p) - (y' - p) f_{y'}(x, y, p) \geq 0 . \tag{12.61}$$

That inequality can be rewritten, in terms of the Weierstrass excess function, as

$$E(x, y, p, y') \geq 0 . \tag{12.62}$$

This inequality should look familiar; it was one of the conditions for a strong minimum from the last section.

We still need to characterize $S(x, y)$ and $p(x, y)$ more fully. Let us introduce the new variable

$$z \equiv \frac{\partial f}{\partial y'}(x, y, p) . \tag{12.63}$$

We may now solve for $p$ in terms of $z$ and introduce the Hamiltonian function

$$H(x, y, z) = z\,p - f(x, y, p) . \tag{12.64}$$

(This is the usual Legendre transformation that we first saw in Chapter 4, but with $z$ now playing the role of the canonical momentum.) It now follows, from Carathéodory's equations, that $S(x, y)$ must satisfy the partial differential equation

$$\frac{\partial S}{\partial x} + H\left(x, y, \frac{\partial S}{\partial y}\right) = 0 . \tag{12.65}$$

Equation (12.65) is known as the *Hamilton–Jacobi equation.*

If we can solve the Hamilton–Jacobi equation for $S(x, y)$, then our canonical momentum is, in light of equations (12.59) and (12.63),

$$z = \frac{\partial S}{\partial y}. \tag{12.66}$$

This variable determines the slope $p(x, y)$ of our field,

$$y' = p(x, y) = \frac{\partial H}{\partial z}\left(x, y, \frac{\partial S}{\partial y}\right). \tag{12.67}$$

We can integrate this slope, subject to boundary conditions (12.48), to minimize our functional and solve our problem.

Taking the ordinary derivative of equation (12.66) gives

$$z' = \frac{\partial^2 S}{\partial x \partial y} + \frac{\partial^2 S}{\partial y^2}\, y'. \tag{12.68}$$

Taking the partial derivative of the Hamilton–Jacobi equation with respect to $y$ gives

$$\frac{\partial^2 S}{\partial x \partial y} + \frac{\partial H}{\partial y} + \frac{\partial H}{\partial z}\frac{\partial^2 S}{\partial y^2} = 0. \tag{12.69}$$

Combining these two equations gives

$$z' = -\frac{\partial H}{\partial y}\left(x, y, \frac{\partial S}{\partial y}\right). \tag{12.70}$$

Equation (12.67) and (12.70) are the canonical or Hamiltonian equations. The solution generated using the Hamilton–Jacobi equation is thus an extremal, a solution of the Euler–Lagrange equation. It is usually easiest to integrate the Euler–Lagrange equation directly, but the Hamilton–Jacobi equation provides us, in some instances, with an alternative means of obtaining this solution.

**Example 12.5.**

Consider the problem of minimizing the distance,

$$J[y] = \int_0^b \sqrt{1 + y'^2}\, dx, \tag{12.71}$$

between the origin and the point $(b, y_b)$. That is, we wish to minimize our functional subject to the boundary conditions

$$y(0) = 0 \quad \text{and} \quad y(b) = y_b. \tag{12.72}$$

For this problem, we begin by introducing the new variable

$$z = \frac{\partial f}{\partial y'}(x, y, p) = \frac{p}{\sqrt{1+p^2}} \tag{12.73}$$

so that

$$p = \frac{z}{\sqrt{1-z^2}}. \tag{12.74}$$

Our Hamiltonian now takes the form

$$\begin{aligned} H(x, y, z) &= z\,p - f(x, y, p) \\ &= \frac{z^2}{\sqrt{1-z^2}} - \sqrt{1 + \frac{z^2}{1-z^2}} \\ &= -\sqrt{1-z^2} \end{aligned} \tag{12.75}$$

and our Hamilton–Jacobi equation is just

$$\frac{\partial S}{\partial x} - \sqrt{1 - \left(\frac{\partial S}{\partial y}\right)^2} = 0 \tag{12.76}$$

or

$$\left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{\partial S}{\partial y}\right)^2 = 1. \tag{12.77}$$

Let us look for a solution that satisfies

$$\left(\frac{\partial S}{\partial x}\right)^2 = \alpha\,, \quad \left(\frac{\partial S}{\partial y}\right)^2 = 1 - \alpha\,. \tag{12.78}$$

Given these equations, it quickly follows that

$$\frac{\partial S}{\partial x} = \sqrt{\alpha} \tag{12.79}$$

so that

$$S(x, y) = \sqrt{\alpha}x + g(y)\,. \tag{12.80}$$

Likewise,

$$\frac{\partial S}{\partial y} = \sqrt{1-\alpha} \tag{12.81}$$

so that

$$S(x, y) = \sqrt{1-\alpha}\,y + h(x)\,. \tag{12.82}$$

Comparing these two solutions, we see that

$$S(x, y) = \sqrt{\alpha}\,x + \sqrt{1-\alpha}\,y + \beta\,. \tag{12.83}$$

Because of equation (12.58), $S(x, y)$, for this problem, is just the distance between the origin and the point $(x, y)$. We thus want a solution to our Hamilton–Jacobi equation that satisfies

$$S(x,0) \;=\; |x|\,, \quad S(0,y) \;=\; |y|\,. \tag{12.84}$$

Sadly, there are no values of $\alpha$ and $\beta$ that will make solution (12.83) satisfy these conditions. Fortunately, there is another solution, a singular integral, lurking nearby.

To obtain this singular integral, we must take the envelope of solution (12.83) by eliminating $\alpha$ between equation (12.83) and the equation

$$\frac{\partial S}{\partial \alpha} \;=\; \frac{x}{2\sqrt{\alpha}} - \frac{y}{2\sqrt{1-\alpha}} \;=\; 0\,. \tag{12.85}$$

This last equation implies that

$$\alpha \;=\; \frac{x^2}{x^2+y^2} \tag{12.86}$$

and, if we substitute this expression for $\alpha$ into solution (12.83), we find that

$$S(x,y) \;=\; \sqrt{x^2+y^2} + \beta\,. \tag{12.87}$$

For $\beta = 0$, we now have no trouble satisfying equations (12.84). For this problem, $S(x, y)$ is, unsurprisingly, the Euclidean distance.

It now quickly follows that

$$z \;=\; \frac{\partial S}{\partial y} \;=\; \frac{y}{\sqrt{x^2+y^2}} \tag{12.88}$$

and, by equation (12.73), that

$$p \;=\; \frac{z}{\sqrt{1-z^2}} \;=\; \frac{y}{x}\,. \tag{12.89}$$

If we integrate the slope of our field,

$$\frac{dy}{dx} = p(x, y) = \frac{y}{x}, \tag{12.90}$$

subject to boundary conditions (12.72), we find, again unsurprisingly, that our solution is the straight line

$$y(x) = \frac{y_b}{b}x. \tag{12.91}$$

Carathéodory's method of equivalent variational problems quickly leads to the Euler–Lagrange equation and to Weierstrass's condition and is the starting point for many modern investigations of the calculus of variations.

## 12.6. Recommended reading

Thiele (1997) and Fraser (2009) provide useful historical surveys of sufficient conditions and of field theory in the calculus of variations.

David Hilbert introduced his invariant integral as part of Problem 23 of his famous lecture at the International Congress of Mathematics in Paris in 1900. Osgood (1901) wrote a fine summary of Hilbert's method a short time later. See also Hedrick (1902). Hilbert's lecture was soon translated into English (Hilbert, 1902) and still makes for fascinating reading.

Carathéodory's method of equivalent variational problems can, of course, be found in his well-known book (Carathéodory, 1935, 2002). Maurin (1997) provides an exceptionally thorough overview of the royal road. The history of Carathéodory's contribution is discussed in Pesch and Bulirsch (1994), Thiele (1997), and Fraser (2009). Carlson (2002) compares Carathéodory's method with that of Leitmann.

Anderson and Arthurs (1999) show how to solve the Hamilton–Jacobi equation for the brachistochrone problem. For more on the Hamilton–Jacobi equation and the calculus of variations, see Rund (1966).

Carathéodory's method and the Hamilton–Jacobi equation provide a bridge between the calculus of variations and its successors, dynamic programming and optimal control theory. See Dreyfus (1965) for more on the connection between the calculus of variations and dynamic programming and Snow (1967) and Pesch and Bulirsch (1994) on the connection between Carathéodory's method and optimal control theory.

## 12.7. Exercises

**12.7.1. Hamilton–Jacobi equation for the brachistochrone.** Consider an integral of the form

$$J[y] \;=\; \int_a^b \sqrt{\frac{1+y'^2}{y_a - y}}\; dx \tag{12.92}$$

and show that it leads to the Hamilton–Jacobi equation

$$\left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{\partial S}{\partial y}\right)^2 \;=\; \frac{1}{y}\,. \tag{12.93}$$

**12.7.2. Weak and strong extrema.** Find and classify the extrema of the following functionals (Elsgolc, 1961):

(a)

$$J[y] \;=\; \int_0^2 (xy' + y'^2)\; dx\,, \tag{12.94}$$

$$y(0) \;=\; 1\,, \quad y(2) \;=\; 0\,. \tag{12.95}$$

(b)

$$J[y] \;=\; \int_0^a (y'^2 + 2yy' - 16y^2)\; dx\,, \tag{12.96}$$

$$a > 0\,, \quad y(0) \;=\; 0\,, \quad y(a) \;=\; 0\,. \tag{12.97}$$

(c)

$$J[y] \;=\; \int_{-1}^2 y'(1 + x^2 y')\, dx\,, \tag{12.98}$$

$$y(-1) \;=\; 1\,, \quad y(2) \;=\; 4\,. \tag{12.99}$$

(d)

$$J[y] \;=\; \int_{1}^{2} y'(1+x^2y')\;dx\,, \tag{12.100}$$

$$y(1) \;=\; 3\,, \quad y(2) \;=\; 5\,. \tag{12.101}$$

(e)

$$J[y] \;=\; \int_{-1}^{2} y'(1+x^2y')\;dx\,, \tag{12.102}$$

$$y(-1) \;=\; y(2) \;=\; 1\,. \tag{12.103}$$

(f)

$$J[y] \;=\; \int_{0}^{\pi/4} (4y^2 - {y'}^2 + 8y)\;dx\,, \tag{12.104}$$

$$y(0) \;=\; -1\,, \quad y(\pi/4) \;=\; 0\,. \tag{12.105}$$

(g)

$$J[y] \;=\; \int_{1}^{2} (x^2{y'}^2 + 12y^2)\;dx\,, \tag{12.106}$$

$$y(1) \;=\; 1\,, \quad y(2) \;=\; 8\,. \tag{12.107}$$

(h)

$$J[y] \;=\; \int_{x_0}^{x_1} \frac{1+y^2}{{y'}^2}\;dx\,, \tag{12.108}$$

$$y(x_0) \;=\; y_0\,, \quad y(x_1) \;=\; y_1\,. \tag{12.109}$$

(i)

$$J[y] \;=\; \int_{0}^{1} ({y'}^2 + y^2 + 2y\,e^{2x})\;dx\,, \tag{12.110}$$

$$y(0) \;=\; \frac{1}{3}\,, \quad y(1) \;=\; \frac{1}{3}\,e^2\,. \tag{12.111}$$

(j)

$$J[y] \;=\; \int\limits_{0}^{\pi/4} (y^2 - {y'}^2 + 6\,y\,\sin\,2x)\;dx\,, \tag{12.112}$$

$$y(0) \;=\; 0\,, \quad y(\pi/4) \;=\; 1\,. \tag{12.113}$$

(k)

$$J[y] \;=\; \int\limits_{0}^{x_1} \frac{1}{y'}\;dx\,, \tag{12.114}$$

$$y(0) \;=\; 0\,, \quad y(x_1) \;=\; y_1\,, \quad x_1 \;>\; 0\,, \quad y_1 \;>\; 0\,. \tag{12.115}$$

(l)

$$J[y] \;=\; \int\limits_{0}^{x_1} \frac{1}{{y'}^2}\;dx\,, \tag{12.116}$$

$$y(0) \;=\; 0\,, \quad y(x_1) \;=\; y_1\,, \quad x_1 \;>\; 0\,, \quad y_1 \;>\; 0\,. \tag{12.117}$$

(m)

$$J[y] \;=\; \int\limits_{1}^{2} \frac{x^3}{{y'}^2}\;dx\,, \tag{12.118}$$

$$y(1) \;=\; 1\,, \quad y(2) \;=\; 4\,. \tag{12.119}$$

(n)

$$J[y] \;=\; \int\limits_{1}^{3} (12xy + {y'}^2)\;dx\,, \tag{12.120}$$

$$y(1) \;=\; 0\,, \quad y(3) \;=\; 26\,. \tag{12.121}$$

# Bibliography


N. I. Akhiezer. *The Calculus of Variations*. Blaisdell Publishing Company, New York, New York, USA, 1962.

N. I. Akhiezer. *The Calculus of Variations*. Harwood Academic Publishers, Chur, Switzerland, 1988.

L. D. Akulenko. The brachistochrone problem for a disc. *Journal of Applied Mathematics and Mechanics*, 73:371–378, 2009.

N. Anderson and A. M. Arthurs. Hamilton–Jacobi results for the brachistochrone. *European Journal of Physics*, 20:101–104, 1999.

P. K. Aravind. Simplified approach to brachistochrone problems. *American Journal of Physics*, 49:884–886, 1981.

J. L. M. Barbosa and A. G. Colares. *Minimal Surfaces in* $\mathbb{R}^3$. Springer-Verlag, Berlin, Germany, 1986.

J.-L. Basdevant. *Variational Principles in Physics*. Springer, New York, New York, USA, 2007.

M. Ben Amar, P. P. da Silva, N. Limodin, A. Langlois, M. Brazovskaia, C. Even, I. V. Chikina, and P. Pieranski. Stability and vibrations of catenoid-shaped smectic films. *European Physical Journal B*, 3:197–202, 1998.

M. Berger. *A Panoramic View of Riemannian Geometry*. Springer, Berlin, Germany, 2003.

Ja. Bernoulli. Solutio problematum fraternorum, peculiari programmate cal. Jan. 1697 Groningæ, nec non Actorum Lips. mense Jun. & Dec. 1696, & Febr. 1697 propositorum: una cum propositione reciproca aliorum. *Acta Eruditorum*, 16:211–217, 1697.

Jo. Bernoulli. Problema novum ad cujus solutionem mathematici invitantur. *Acta Eruditorum*, 15:269, 1696.

Jo. Bernoulli. Curvatura radii in diaphanis non uniformibus, solutioque problematis a se in Actis 1696, p. 269, propositi, de invenienda linea brachystochrona, id est, in qua grave a dato puncto ad datum punctum brevissimo tempore decurrit, & de curva synchrona seu readiorum unda construenda. *Acta Eruditorum*, 16: 206–211, 1697a.

Jo. Bernoulli. Problème à résoudre. *Journal des Sçavans*, 26 August 1697:394–396, 1697b.

Jo. Bernoulli. Remarques sur ce qu'on a donné jusqu'ici de solutions des problémes sur des isopérimetrés; avec une nouvelle méthode courte & facile de les resoudre sans calcul, laquelle s'étend aussi à d'austres problêmes qui ont raport à ceux-là. *Mémoires de l'Académie Royale des Sciences, Paris*, 2:235–269, 1718.

G. A. Bliss. The geodesic lines on the anchor ring. *Annals of Mathematics*, 4:1–21, 1902.

G. A. Bliss. A note on functions of lines. *Proceedings of the National Academy of Sciences of the United States of America*, 1:173–177, 1915.

G. A. Bliss. Jacobi's condition for problems of the calculus of variations in parametric form. *Transactions of the American Mathematical Society*, 17:195–206, 1916.

G. A. Bliss. The problem of Lagrange in the calculus of variations. *American Journal of Mathematics*, 52:673–744, 1930.

H. Boerner. Carathéodory's eingang zur variationsrechnung. *Jahresbericht der Deutsche Mathematiker-Vereinigung*, 56:31–58, 1953.

O. Bolza. The determination of the conjugate points for discontinuous solutions in the calculus of variations. *American Journal of Mathematics*, 30:209–221, 1908.

O. Bolza. *Lectures on the Calculus of Variations.* Chelsea Publishing Company, New York, New York, USA, 1973.

M. Born and E. Wolf. *Principles of Optics.* Cambridge University Press, Cambridge, UK, 1999.

U. Brechtken-Manderscheid. *Introduction to the Calculus of Variations.* Chapman & Hall, London, UK, 1991.

W. E. Brooke and H. B. Wilcox. *Engineering Mechanics.* Ginn and Company, Boston, Massachusetts, USA, 1929.

G. Buttazzo and B. Kawohl. On Newton's problem of minimal resistance. *Mathematical Intelligencer*, 15:7–12, 1993.

R. S. Capon. Hamilton's principle in relation to nonholonomic mechanical systems. *Quarterly Journal of Mechanics and Applied Mathematics*, 5:472–480, 1952.

C. Carathéodory. *Variationsrechnung und Partielle Differentialgleichungen Erster Ordnung.* Teubner, Leipzig, Germany, 1935.

C. Carathéodory. The beginnings of research in the calculus of variations. *Osiris*, 3:224–240, 1937.

C. Carathéodory. *Calculus of Variations and the Partial Differential Equations of the First Order.* AMS Chelsea Publishing, Providence, Rhode Island, USA, 2002.

D. A. Carlson. An observation on two methods of obtaining solutions to variational problems. *Journal of Optimization Theory and Applications*, 114:345–361, 2002.

R. Chander. Gravitational fields whose brachistochrones and isochrones are identical curves. *American Journal of Physics*, 45: 848–850, 1977.

Y.-J. Chen and P. H. Steen. Dynamics of inviscid capillary breakup: collapse and pinchoff of a film bridge. *Journal of Fluid Mechanics*, 341:245–267, 1997.

A. C. Clairaut. Détermination géométrique de la perpendiculaire à la méridienne tracée par M. Cassini; avec plusieurs méthodes d'en tirer la grandeur et la figure de la terre. *Histoire de l'Académie Royale des Sciences avec les Mémoires de Mathématique et de Physique*, 1735:406–416, 1733.

J. C. Clegg. *Calculus of Variations*. Oliver & Boyd, Edinburgh, UK, 1968.

P. W. Cooper. Through the earth in forty minutes. *American Journal of Physics*, 34:68–70, 1966a.

P. W. Cooper. Further commentary on "Through the Earth in Forty Minutes". *American Journal of Physics*, 34:703–704, 1966b.

C. Criado and N. Alamo. Solving the brachistochrone and other variational problems with soap films. *American Journal of Physics*, 78:1400–1405, 2010.

J. Cruz-Sampedro and M. Tetlalmatzi-Montiel. Minimum resistance in a rare medium. In J. Klapp and A. Medina, editors, *Experimental and Computational Fluid Mechanics*, pages 129–145. Springer, Cham, Switzerland, 2014.

S. A. Cryer and P. H. Steen. Collapse of the soap-film bridge: quasi-static description. *Journal of Colloid and Interface Science*, 154: 276–288, 1992.

L. D'Antonio. "The fabric of the universe is most perfect": Euler's research on elastic curves. In R. E. Bradley, L. A. D'Antonio, and C. E. Sandifer, editors, *Euler at 300: An Appreciation*, pages 239–260. Mathematical Association of America, Washington, D.C., USA, 2007.

G. Della Riccia. On the Lagrange representation of a system of Newton equations. In A. Avez, Z. Blaquière, and A. Marzollo, editors, *Dynamical Systems and Microphysics: Geometry and Mechanics*, pages 281–292. Academic Press, New York, New York, USA, 1982.

H. H. Denman. Remarks on brachistochrone-tautochrone problems. *American Journal of Physics*, 53:224–227, 1985.

A. Dresden. An example of the indicatrix in the calculus of variations. *American Mathematical Monthly*, 14:119–126, 1907.

A. Dresden. The second derivatives of the extremal integral. *Transactions of the American Mathematical Society*, 9:467–486, 1908.

A. Dresden. Five theses on calculus of variations. *Bulletin of the American Mathematical Society*, 38:617–621, 1932.

S. E. Dreyfus. *Dynamic Programming and the Calculus of Variations.* Academic Press, New York, New York, USA, 1965.

P. du Bois-Reymond. Erläuterungen zu den anfangsgründen variationsrechnung. *Mathematische Annalen*, 15:283–314, 1879a.

P. du Bois-Reymond. Fortsetzung der erläuterungen zu den anfangsgründen variationsrechnung. *Mathematische Annalen*, 15:564–576, 1879b.

R. Dugas. *A History of Mechanics.* Dover Publications, New York, New York, USA, 1988.

L. Durand. Stability and oscillations of a soap film: an analytic treatment. *American Journal of Physics*, 49:334–343, 1981.

H.-D. Ebbinghaus. *Ernst Zermelo: An Approach to His Life and Work.* Springer, Berlin, Germany, 2007.

D. G. B. Edelen. Aspects of variational arguments in the theory of elasticity: fact and folklore. *International Journal of Solids and Structures*, 17:729–740, 1981.

L. K. Edwards. High-speed tube transportation. *Scientific American*, 213(2):30–40, 1965.

L. E. Elsgolc. *Calculus of Variations.* Pergamon Press, London, UK, 1961.

G. Erdmann. Ueber unstetige lösungen in der variationsrechnung. *Journal für die Reine und Ungewandte Matehmatik*, 82:21–30, 1877.

G. Erdmann. Zur untersuchung der zweiten variation einfacher intregale. *Zeitschrift für Mathematik und Physik*, 23:362–379, 1878.

H. Erlichson. Johann Bernoulli's brachistochrone solution using Fermat's principle of least time. *European Journal of Physics*, 20: 299–304, 1999.

L. Euler. De linea brevissima in superficie quacunque duo quaelibet puncta jungente. *Commentarii Academiae Scientiarum Petropolitanae*, 3:110–124, 1732.

L. Euler. *Methodus Inveniendi Lineas Curvas Maximi Minimive Proprietate Gaudentes, sive Solutio Problematis Isoperimetrici Latissimo Sensu Accepti.* Bousquet, Lausanne and Geneva, 1744.

J. Evans and M. Rosenquist. “F = ma” optics. *American Journal of Physics*, 54:876–883, 1986.

C. Farina. Bernoulli’s method for relativistic brachistochrones. *Journal of Physics A: Mathematical and General*, 20:L57–L59, 1987.

U. Filobello-Nino, H. Vazquez-Leal, D. Pereyra-Diaz, A. Yildirim, A. Perez-Sesma, R. Castaneda-Sheissa, J. Sanchez-Orea, and C. Hoyos-Reyes. A generalization of the Bernoulli’s methods applied to brachistochrone-like problems. *Applied Mathematics and Computation*, 219:6707–6718, 2013.

M. R. Flannery. The enigma of nonholonomic constraints. *American Journal of Physics*, 73:265–272, 2005.

A. T. Fomenko. *The Plateau Problem. Part I: Historical Survey. Part II: The Present State of the Theory.* Gordon and Breach Science Publishers, New York, New York, USA, 1990.

A. T. Fomenko and A. A. Tuzhilin. *Elements of the Geometry and Topology of Minimal Surfaces in Three-Dimensional Space.* American Mathematical Society, Providence, Rhode Island, USA, 1991.

G. W. Forbes. On variational problems in parametric form. *American Journal of Physics*, 59:1130–1140, 1991.

M. J. Forray. *Variational Calculus in Science and Engineering.* McGraw-Hill Book Company, New York, New York, USA, 1968.

A. R. Forsyth. *Calculus of Variations.* Cambridge University Press, Cambridge, UK, 1927.

C. Fox. *An Introduction to the Calculus of Variations.* Oxford University Press, London, UK, 1950.

C. Fraser. J. L. Lagrange’s changing approach to the foundations of the calculus of variations. *Archive for History of Exact Sciences*, 32:151–191, 1985.

C. Fraser. The calculus of variations: a historical survey. In H. N. Jahnke, editor, *A History of Analysis*, pages 355–383. American Mathematical Society, Providence, Rhode Island, USA, 2003.

C. G. Fraser. Mathematical technique and physical conception in Euler’s investigation of the elastica. *Centaurus*, 34:211–246, 1991.

C. G. Fraser. Isoperimetric problems in the variational calculus of Euler and Lagrange. *Historia Mathematica*, 19:4–23, 1992.

C. G. Fraser. The origins of Euler's variational calculus. *Archive for History of Exact Sciences*, 47:103–141, 1994.

C. G. Fraser. Leonhard Euler, book on the calculus of variations (1744). In I. Grattan-Guinness, editor, *Landmark Writings in Western Mathematics 1640–1940*, pages 168–180. Elsevier, Amsterdam, Netherlands, 2005a.

C. G. Fraser. Joseph Louis Lagrange, *Théorie des Fonctions Analytiques*, first edition (1797). In I. Grattan-Guinness, editor, *Landmark Writings in Western Mathematics 1640–1940*, pages 258–276. Elsevier, Amsterdam, Netherlands, 2005b.

C. G. Fraser. Sufficient conditions, fields and the calculus of variations. *Historia Mathematica*, 36:420–427, 2009.

I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, USA, 1963.

R. D. Gillette and D. C. Dyson. Stability of fluid interfaces of revolution between equal solid circular plates. *Chemical Engineering Journal*, 2:44–54, 1971.

B. Goldschmidt. *Determinatio superficiei minimae rotatione curvae data duo puncta jungentis circa datum axem ortae*. Göttingen Prize Essay, Göttingen, Germany, 1831.

H. Goldstein. *Classical Mechanics*. Addison-Wesley Publishing Company, Reading, Massachusetts, USA, 1980.

H. F. Goldstein and C. M. Bender. Relativistic brachistochrone. *Journal of Mathematical Physics*, 27:507–511, 1986.

H. H. Goldstine. *A History of the Calculus of Variations from the 17th Through the 19th Century*. Springer-Verlag, New York, New York, USA, 1980.

V. G. A. Goss. The history of the planar elastica: insights into mechanics and scientific method. *Science and Education*, 18:1057–1082, 2009.

L. M. Graves. Discontinuous solutions in the calculus of variations. *Bulletin of the American Mathematical Society*, 36:831–846, 1930a.

L. M. Graves. Discontinuous solutions in space problems of the calculus of variations. *American Journal of Mathematics*, 52:1–28, 1930b.

L. M. Graves. A proof of the Weierstrass condition in the calculus of variations. *American Mathematical Monthly*, 41:502–504, 1934.

C. G. Gray and E. F. Taylor. When action is not least. *American Journal of Physics*, 75:434–458, 2007.

H. Haichang. *Variational Principles of Theory of Elasticity with Applications.* Gordon and Breach, New York, New York, USA, 1984.

E. P. Hamilton. A new definition of variational derivative. *Bulletin of the Australian Mathematical Society*, 22:205–210, 1980.

E. P. Hamilton and M. Z. Nashed. Global and local variational derivatives and integral representations of Gâteaux differentials. *Journal of Functional Analysis*, 49:128–144, 1982.

E. P. Hamilton and M. Z. Nashed. Variational derivatives in function spaces. In V. Lakshmikantham, editor, *World Congress of Nonlinear Analysts '92: Proceedings of the First World Congress of Nonlinear Analysts, Tampa, Florida, August 19–26, 1992*, pages 527–536. Walter de Gruyter, Berlin, Germany, 1995.

W. R. Hamilton. On a general method in dynamics; by which the study of the motions of all free systems of attracting or repelling points is reduced to the search and differentiation of one central relation, or characteristic function. *Philosophical Transactions of the Royal Society of London*, 124:247–308, 1834.

W. R. Hamilton. Second essay on a general method in dynamics. *Philosophical Transactions of the Royal Society of London*, 125: 95–144, 1835.

H. Hancock. *Lectures on the Calculus of Variations.* Cincinatti University Press, Cincinatti, Ohio, USA, 1904.

E. R. Hedrick. On the sufficient conditions in the calculus of variations. *Bulletin of the American Mathematical Society*, 9:11–24, 1902.

H. Hertz. *The Principles of Mechanics: Presented in a New Form.* MacMillan and Company, New York, New York, USA, 1899.

M. R. Hestenes. A note on the Jacobi condition for parametric problems in the calculus of variations. *Bulletin of the American Mathematical Society*, 40:297–302, 1934.

J. Heyman. *Elements of the Theory of Structures*. Cambridge University Press, Cambridge, UK, 1996.

J. Heyman. *Structural Analysis: A Historical Approach*. Cambridge University Press, Cambridge, UK, 1998.

D. Hilbert. Mathematical problems: lecture delivered before the International Congress of Mathematicians at Paris in 1900. *Bulletin of the American Mathematical Society*, 8:437–479, 1902.

K. A. Hoffman. Stability results for constrained calculus of variations problems: an analysis of the twisted elastic loop. *Proceedings of the Royal Society of London A*, 461:1357–1381, 2005.

K. A. Hoffman, R. S. Manning, and R. C. Paffenroth. Calculation of the stability index in parameter-dependent calculus of variations problems: buckling of a twisted elastic strut. *SIAM Journal of Applied Dynamical Systems*, 1:115–145, 2002.

D. Horstmann, B. Kawohl, and P. Villaggio. Newton's aerodynamic problem in the presence of friction. *NoDEA. Nonlinear Differential Equations and Applications*, 9:295–307, 2002.

A. Huke. *An historical and critical study of the fundamental lemma in the calculus of variations*. The University of Chicago Press, Chicago, Illinois, USA, 1931.

J. E. Hurtado. Paul Bunyan's brachistochrone and tautochrone. *Journal of the Astronautical Sciences*, 48:207–224, 2000.

C. Huygens. *Horologium Oscillatorium Sive de Motu Pendulorum ad Horologia Aptato Demonstrationes Geometricae*. Apud F. Muguet, Paris, France, 1673.

C. Huygens. *Christiaan Huygens' The Pendulum Clock or Geometrical Demonstrations Concerning the Motion of Pendula as Applied to Clocks*. Iowa State University Press, Ames, Iowa, USA, 1986.

M. de Icaza Herrera. Galileo, Bernoulli, Leibniz and Newton around the brachistochrone problem. *Revista Mexicana de Física*, 40:459–475, 1994.

C. Isenberg. *The Science of Soap Films and Soap Bubbles.* Dover Publications, Inc., Mineola, New York, USA, 1992.

C. G. J. Jacobi. Zur theorie der variations-rechnung und der differential-gleichungen. *Journal für die Reine und Angewandte Mathematik*, 17:68–82, 1837.

C. G. J. Jacobi. Note von der geodätischen linie auf einem ellipsoid und der verschiedenen anwendungen einer merkwürdigen analytischen substitution. *Journal für die Reine und Angewandte Mathematik*, 19:309–313, 1839.

H. Jeffreys. What is Hamilton's principle? *Quarterly Journal of Mechanics and Applied Mathematics*, 7:335–337, 1954.

J. N. Joglekar and W. K. Tham. Exploring the action landscape via trial world-lines. *European Journal of Physics*, 32:129–138, 2011.

W. Johnson. Isaac Todhunter (1820–1884): textbook writer, scholar, coach and historian of science. *International Journal of Mechanical Sciences*, 38:1231–1270, 1996.

S. G. Kamath. Relativistic tautochrone. *Journal of Mathematical Physics*, 33:934–940, 1992.

J. C. Kimball and H. Story. Fermat's principle, Huygens' principle, Hamilton's optics and sailing strategy. *European Journal of Physics*, 19:15–24, 1998.

W. S. Kimball. *Calculus of Variations by Parallel Displacement.* Butterworths Scientific Publications, London, UK, 1952.

P. G. Kirmser. An example of the need for adequate references. *American Journal of Physics*, 34:701, 1966.

M. Kline. *Mathematical Thought from Ancient to Modern Times.* Oxford University Press, New York, New York, USA, 1972.

A. Kneser. *Lehrbuch der Variationsrechnung.* F. Vieweg und Sohn, Braunschweig, Germany, 1900.

A. N. Kolmogorov and A. P. Yushkevich. *Mathematics of the 19th Century: Function Theory According to Chebyshev, Ordinary Differential Equations, Calculus of Variations, Theory of Finite Differences.* Birkhauser Verlag, Basel, Switzerland, 1998.

T. Lachand-Robert and E. Oudet. Minimizing within convex bodies using a convex hull method. *SIAM Journal of Optimization*, 16: 368–379, 2005.

J. L. Lagrange. Essai d'une nouvelle méthode pour déterminer les maxima et les minima des formules intégrales indéfinies. *Miscellanea Taurinensia*, 2(1760/61):173–195, 1762.

J. L. Lagrange. *Théorie des Fonctions Analytiques.* L'Imprimerie de la République, Paris, France, 1797.

V. Lakshminarayanan, A. K. Ghatak, and K. Thyagarajan. *Lagrangian Optics.* Kluwer Academic Publishers, Boston, Massachusetts, USA, 2002.

C. Lanczos. *The Variational Principles of Mechanics.* University of Toronto Press, Toronto, Ontario, Canada, 1974.

L. J. Laslett. Trajectory for minimum transit time through the earth. *American Journal of Physics*, 34:702–703, 1966.

L. P. Lebedev and M. J. Cloud. *The Calculus of Variations and Functional Analysis with Optimal Control and Applications in Mechanics.* World Scientific, Singapore, 2003.

H. Lebesgue. Intégrale, longueur, aire. *Annali di Matematica Pura ed Applicata*, 7:231–359, 1902.

A. M. Legendre. Mémoires sur la mainère de distinguer les maxima des minima sans le calcul des variations. *Histoire de l'Académie Royale des Sciences avec les Mémoires de Mathématique et de Physique*, (1786):7–37, 1788.

V. P. Legeza. Brachistochrone for a rolling cylinder. *Mechanics of Solids*, 45:27–33, 2010.

G. W. Leibniz. Communicatio suæ pariter, duarumque alienarum ad edendum sibi primum a Dn. Jo. Bernoullio, deinde a Dn. Marchione Hospitalio communicatarum solutionum problematis curvæ celerrimi descensus a Dn. Jo. Bernoullio geometris publice propositi, una cum solutione sua problematis alterius ab eodem postea propositi. *Acta Eruditorum*, 16:201–205, 1697.

A. W. Leissa. The historical bases of the Rayleigh and Ritz methods. *Journal of Sound and Vibration*, 287:961–978, 2005.

D. S. Lemons. *Perfect Form: Variational Principles, Methods, and Applications in Elementary Physics.* Princeton University Press, Princeton, New Jersey, USA, 1997.

G. F. A. de l'Hôpital. Solutio problematis de linea celerrimi descensus. *Acta Eruditorum*, 16:217–218, 1697.

L. Lindelöf and M. Moigno. *Leçons de Calcul des Variations.* Mallet-Bachelier, Paris, France, 1861.

B. J. Lowry and P. H. Steen. Capillary surfaces: stability from families of equilibria with applications to the liquid bridge. *Proceedings of the Royal Society of London A*, 449:411–439, 1995.

J. Lützen. *Mechanistic Images in Geometric Form: Heinrich Hertz's Principles of Mechanics.* Oxford University Press, New York, New York, USA, 2005.

R. L. Mallett. Comments on "Through the Earth in Forty Minutes". *American Journal of Physics*, 34:702, 1966.

R. S. Manning. Conjugate points revisited and Neumann-Neumann problems. *SIAM Review*, 51:193–212, 2009.

R. S. Manning, K. A. Rodgers, and J. H. Maddocks. Isoperimetric conjugate points with application to the stability of DNA minicircles. *Proceedings of the Royal Society of London A*, 454:3047–3074, 1998.

E. W. Marchand. *Gradient Index Optics.* Academic Press, New York, New York, USA, 1978.

K. Maurin. *The Riemann Legacy: Riemannian Ideas in Mathematics and Physics.* Kluwer Academic Publishers, Dordrecht, Netherlands, 1997.

J. M. McKinley. Brachistochrones, tautochrones, evolutes, and tessellations. *American Journal of Physics*, 47:81–86, 1979.

A. Mercier. *Analytical and Canonical Formalism in Physics.* North-Holland Publishing Company, Amsterdam, Netherlands, 1959.

A. S. Merrill. An isoperimetric problem with variable end-points. *American Journal of Mathematics*, 41:60–78, 1919.

S. Mertens and S. Mingramm. Brachistochrones with loose ends. *European Journal of Physics*, 29:1191–1199, 2008.

J. B. M. C. Meusnier. Mémoire sur la courbure des surfaces. *Mémoires de Mathématique et de Physique Présentés à l'Académie Royal des Sciences, par Divers Sçavans, et Lus dans ses Assemblées*, 10:447–510, 1785.

A. Miele. *Theory of Optimum Aerodynamic Shapes: Extremal Problems in the Aerodynamics of Supersonic, Hypersonic, and Free-Molecular Flows*. Academic Press, New York, New York, USA, 1965.

J. Milnor. *Morse Theory*. Princeton University Press, Princeton, New Jersey, USA, 1963.

B. L. Moiseiwitsch. *Variational Principles*. John Wiley & Sons, London, UK, 1966.

M. Morse. *The Calculus of Variations in the Large*. American Mathematical Society, Providence, Rhode Island, USA, 1934.

F. Müller and R. Stannarius. Collapse of catenoid-shaped smectic films. *Europhysics Letters*, 76:1102–1108, 2006.

C. E. Mungan and T. C. Lipscombe. Complementary curves of descent. *European Journal of Physics*, 34:59–65, 2013.

M. Nakane and C. G. Fraser. The early history of Hamilton-Jacobi dynamics 1834-1837. *Centaurus*, 44:161–227, 2002.

I. Newton. Epistola missa ad prænobilem virum D. Carolum Montague armigerum, scaccarii regii apud Anglos cancellarium, et societatis regiæ præsidem: in qua solvuntur duo problemata mathematica a Johanne Bernoullio mathematico celeberrimo proposita. *Philosophical Transactions*, 19:388–389, 1695–7.

I. Newton. Epistola missa ad prænobilem virum D. Carolum Montague armigerum, scaccarii regii apud Anglos cancellarium, et societatis regiæ præsidem: in qua solvuntur duo problemata mathematica a Johanne Bernoullio mathematico celeberrimo proposita. *Acta Eruditorum*, 16:223–224, 1697.

J. C. C. Nitsche. *Lectures on Minimal Surfaces, Volume I*. Cambridge University Press, Cambridge, UK, 1989.

T. Okoshi. *Optical Fibers*. Academic Press, New York, New York, USA, 1982.

W. A. Oldfather, C. A. Ellis, and D. M. Brown. Leonhard Euler's elastic curves. *Isis*, 20:72–160, 1933.

J. Oprea. *The Mathematics of Soap Films: Explorations with Maple.* American Mathematical Society, Providence, Rhode Island, USA, 2000.

J. Oprea. *Differential Geometry and Its Applications.* Mathematical Association of America, Washington, D.C., USA, 2007.

O. M. O'Reilly and D. M. Peters. On stability analyses of three classical buckling problems for the elastic strut. *Journal of Elasticity*, 105:117–136, 2011.

W. F. Osgood. Sufficient conditions in the calculus of variations. *Annals of Mathematics*, 2:105–129, 1901.

A. S. Parnovsky. Some generalisations of brachistochrone problem. *Acta Physica Polonica A*, 93:S55–S63, 1998.

L. A. Pars. Variation principles in dynamics. *Quarterly Journal of Mechanics and Applied Mathematics*, 7:338–351, 1954.

L. A. Pars. *An Introduction to the Calculus of Variations.* Heinemann, London, UK, 1962.

J. Patel. Solution of the second-order nonlinear differential equation for the terrestrial brachistochrone. *American Journal of Physics*, 35:436–437, 1967.

N. M. Patrikalakis and T. Maekawa. *Shape Interrogation for Computer Aided Design and Manufacturing.* Springer, Berlin, Germany, 2002.

H. J. Pesch and R. Bulirsch. The maximum principle, Bellman equation, and Carathéodory's work. *Journal of Optimization Theory and Applications*, 80:199–225, 1994.

I. P. Petrov. *Variational Methods in Optimum Control Theory.* Academic Press, New York, New York, USA, 1968.

F. M. Phelps III, F. M. Phelps IV, B. Zorn, and J. Gormley, An experimental study of the brachistochrone, *European Journal of Physics*, 3:1–4, 1982.

J. Plateau. *Statique Expérimenale et Théorique des Liquides Soumis aux Seules Forces Moléculaires.* Gautheir-Villars, Paris, France, 1873.

J. E. Prussing. Brachistochrone-tautochrone problem in a homogeneous sphere. *American Journal of Physics*, 44:304–305, 1976.

S. S. Rao. *Vibration of Continuous Systems.* John Wiley & Sons, Hoboken, New Jersey, USA, 2007.

J. N. Reddy. *Energy Principles and Variational Methods in Applied Mechanics.* John Wiley & Sons, Hoboken, New Jersey, USA, 2002.

N. D. Robinson and P. H. Steen. Observations of singularity formation during the capillary collapse and bubble pinch-off of a soap film bridge. *Journal of Colloid and Interface Science*, 241:448–458, 2001.

E. Rodgers. Brachistochrone and tautochrone curves for rolling bodies. *American Journal of Physics*, 14:249–252, 1946.

R. P. Roess and G. Sansone. *The Wheels That Drove New York: A History of the New York City Transit System.* Springer, Berlin, Germany, 2013.

V. V. Rumiantsev. On integral principles for nonholonomic systems. *Journal of Applied Mathematics and Mechanics*, 46:1–8, 1982.

H. Rund. *The Hamilton–Jacobi theory in the Calculus of Variations.* D. Van Nostrand Company, London, UK, 1966.

H. Sagan. *Introduction to the Calculus of Variations.* McGraw-Hill Book Company, New York, New York, USA, 1969.

W. Sarlet. Symmetries, first integrals and the inverse problem of Lagrangian mechanics. *Journal of Physics A: Mathematical and General*, 14:2227–2238, 1981.

C. J. Silva and D. F. M. Torres. Two-dimensional Newton's problem of minimal resistance. *Control and Cybernetics*, 35:965–975, 2006.

B. Singh and R. Kumar. Brachistochrone problem in nonuniform gravity. *Indian Journal of Pure and Applied Mathematics*, 19:575–585, 1988.

M. A. Slawinski. *Waves and Rays in Elastic Continua.* World Scientific, Singapore, 2010.

D. R. Smith. *Variational Methods in Optimization.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, USA, 1974.

D. R. Smith and C. V. Smith. When is Hamilton's prinicple an extremum principle? *AIAA Journal*, 12:1573–1576, 1974.

J. Sneyd and C. S. Peskin. Computation of geodesic trajectories on tubular surfaces. *SIAM Journal of Scientific and Statistical Computing*, 11:230–241, 1990.

D. R. Snow. Carathéodory–Hamilton–Jacobi theory in optimal control. *Journal of Mathematical Analysis and Applications*, 17:99–118, 1967.

H. L. Stalford and F. E. Garrett. Classical differential geometry solution of the brachistochrone tunnel problem. *Journal of Optimization Theory and Applications*, 80:227–260, 1994.

O. N. Stavroudis. *The Optics of Rays, Wavefronts, and Caustics.* Academic Press, New York, New York, USA, 1972.

O. N. Stavroudis. *The Mathematics of Geometrical and Physical Optics: The k-function and its Ramifications.* Wiley-VCH, Weinheim, Germany, 2006.

E. Stein and K. Weichmann. New insight into optimization and variational problems in the 17th century. *Engineering Computations*, 20:699–724, 2003.

D. J. Struik. Outline of a history of differential geometry: I. *Isis*, 19: 92–120, 1933.

D. J. Struik. *Lectures on Classical Differential Geometry.* Addison-Wesley Publishing Company, Reading, Massachusetts, USA, 1961.

D. J. Struik. *A Source Book of Mathematics, 1200–1800.* Harvard University Press, Cambridge, Massachusetts, USA, 1969.

H. J. Sussmann and J. C. Willems. 300 years of optimal control: from the brachystochrone to the maximum principle. *IEEE Control Systems Magazine*, 17:32–44, 1997.

B. Tabarrok and F. P. J. Rimrott. *Variational Methods and Complementary Formulations in Dynamics.* Kluwer Academic Publishers, Dordrecht, Netherlands, 1994.

G. J. Tee. Isochrones and brachistochrones. *Neural, Parallel and Scientific Computing*, 7:311–341, 1999.

R. Thiele. On some contributions to field theory in the calculus of variations from Beltrami to Carathéodory. *Historia Mathematica*, 24:281–300, 1997.

R. Thiele. Euler and the calculus of variations. In R. E. Bradley and C. E. Sandifer, editors, *Leonhard Euler: Life, Work and Legacy*, pages 235–254. Elsevier, Amsterdam, Netherlands, 2007.

I. Todhunter. *Researches in the Calculus of Variations, Principally in the Theory of Discontinuous Solutions.* McMillan and Company, London and Cambridge, 1871.

I. Todhunter. *A History of the Calculus of Variations During the Nineteenth Century.* Dover Publications, Inc., Mineola, New York, USA, 2005.

C. Truesdell. *The Rational Mechanics of Flexible or Elastic Bodies 1638–1788: Introduction to Leonhardi Euleri Opera Omnia Vol. X et XI Seriei Secundae.* Orell Fussli Turici, Zurich, Switzerland, 1960.

C. Truesdell. The influence of elasticity on analysis: the classic heritage. *Bulletin of the American Mathematical Society*, 9:293–310, 1983.

F. A. Valentine. The problem of Lagrange with differential inequalities as added side contributions. In *Contributions to the Calculus of Variations 1933–1977*, pages 407–448. University of Chicago Press, Chicago, Illinois, USA, 1937.

B. van Brunt. *The Calculus of Variations.* Springer-Verlag, New York, New York, USA, 2004.

G. Venezian. Terrestrial brachistochrone. *American Journal of Physics*, 34:701, 1966.

V. Volterra. Sopra le funzioni che dipendo da altre funzioni. Nota 1, 2, 3. *Atti della Reale Accademia dei Lincei. Serie Quarta. Rendiconti*, 3(Semestre 2):97–105, 141–146, 151–158, 1887.

V. Volterra. *Leçons sur les Fonctions de Lignes.* Gauthier–Villars, Paris, France, 1913.

B. D. Vujanovic and T. M. Atanackovic. *An Introduction to Modern Variational Techniques in Mechanics and Engineering.* Birkhauser, Boston, Massachusetts, USA, 2004.

B. D. Vujanovic and S. E. Jones. *Variational Methods in Nonconservative Phenomena.* Academic Press, San Diego, California, USA, 1989.

D. V. Wallerstein. *A Variational Approach to Structural Analysis.* John Wiley & Sons, New York, New York, USA, 2002.

F. Y. M. Wan. *Introduction to the Calculus of Variations and its Applications.* Chapman & Hall, New York, New York, USA, 1995.

J. K. Whittemore. Lagrange's equation in the calculus of variations and the extension of a theorem of Erdmann. *Annals of Mathematics*, 2:130–136, 1900–1901.

J. Yang, D. G. Stork, and D. Galloway. The rolling unrestrained brachistochrone. *American Journal of Physics*, 55:844–847, 1987.

W. Yourgrau and S. Mandelstam. *Variational Principles in Dynamics and Quantum Theory.* W. B. Saunders, Philadelphia, Pennsylvania, USA, 1968.

R. K. P. Zia, E. F. Redish, and S. R. McKay. Making sense of the Legendre transform. *American Journal of Physics*, 77:614–622, 2009.

# Index

# Selected Published Titles in This Series

72 **Mark Kot,** A First Course in the Calculus of Variations, 2014

71 **Joel Spencer,** Asymptopia, 2014

70 **Lasse Rempe-Gillen and Rebecca Waldecker,** Primality Testing for Beginners, 2014

69 **Mark Levi,** Classical Mechanics with Calculus of Variations and Optimal Control, 2014

68 **Samuel S. Wagstaff, Jr.,** The Joy of Factoring, 2013

67 **Emily H. Moore and Harriet S. Pollatsek,** Difference Sets, 2013

66 **Thomas Garrity, Richard Belshoff, Lynette Boos, Ryan Brown, Carl Lienert, David Murphy, Junalyn Navarra-Madsen, Pedro Poitevin, Shawn Robinson, Brian Snyder, and Caryn Werner,** Algebraic Geometry, 2013

65 **Victor H. Moll,** Numbers and Functions, 2012

64 **A. B. Sossinsky,** Geometries, 2012

63 **María Cristina Pereyra and Lesley A. Ward,** Harmonic Analysis, 2012

62 **Rebecca Weber,** Computability Theory, 2012

61 **Anthony Bonato and Richard J. Nowakowski,** The Game of Cops and Robbers on Graphs, 2011

60 **Richard Evan Schwartz,** Mostly Surfaces, 2011

59 **Pavel Etingof, Oleg Golberg, Sebastian Hensel, Tiankai Liu, Alex Schwendner, Dmitry Vaintrob, and Elena Yudovina,** Introduction to Representation Theory, 2011

58 **Álvaro Lozano-Robledo,** Elliptic Curves, Modular Forms, and Their L-functions, 2011

57 **Charles M. Grinstead, William P. Peterson, and J. Laurie Snell,** Probability Tales, 2011

56 **Julia Garibaldi, Alex Iosevich, and Steven Senger,** The Erdős Distance Problem, 2011

55 **Gregory F. Lawler,** Random Walk and the Heat Equation, 2010

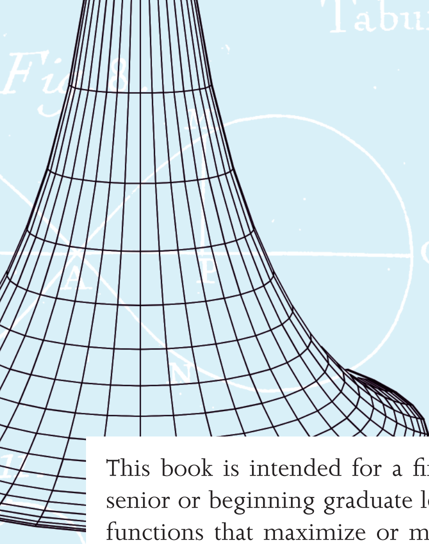
54 **Alex Kasman,** Glimpses of Soliton Theory, 2010

53 **Jiří Matoušek,** Thirty-three Miniatures, 2010

52 **Yakov Pesin and Vaughn Climenhaga,** Lectures on Fractal Geometry and Dynamical Systems, 2009

51 **Richard S. Palais and Robert A. Palais,** Differential Equations, Mechanics, and Computation, 2009

For a complete list of titles in this series, visit the
AMS Bookstore at **www.ams.org/bookstore/stmlseries/**.



This book is intended for a first course in the calculus of variations, at the senior or beginning graduate level. The reader will learn methods for finding functions that maximize or minimize integrals. The text lays out important necessary and sufficient conditions for extrema in historical order, and it illustrates these conditions with numerous worked-out examples from mechanics, optics, geometry, and other fields.

The exposition starts with simple integrals containing a single independent variable, a single dependent variable, and a single derivative, subject to weak variations, but steadily moves on to more advanced topics, including multivariate problems, constrained extrema, homogeneous problems, problems with variable endpoints, broken extremals, strong variations, and sufficiency conditions. Numerous line drawings clarify the mathematics.

Each chapter ends with recommended readings that introduce the student to the relevant scientific literature and with exercises that consolidate understanding.

For additional information and updates on this book, visit
**www.ams.org/bookpages/stml-72**

ISBN: 978-1-4704-1495-5

9 781470 414955

**STML/72**